Advances in Wireless Video Delivery

MINORU ETOH, MEMBER, IEEE, AND TAKESHI YOSHIMURA, MEMBER, IEEE

Invited Paper

This paper reviews practical video delivery technologies, examining existing mobile networks, commercialized or standardized transport, and coding technologies. Compression efficiency, power dissipation, and error control are intrinsic issues in wireless video delivery. Among these issues, error control technologies are evaluated at four layers: 1) layer-1/2 transport; 2) end-to-end transport layer such as TCP/IP or RTP/UDP/IP; 3) error-resilience tool and network adaptation layer) and 4) source coder layer. Layer-1/2 transport tends to provide two distinct conditions: one quasi-error-free, in which upper layer error control technologies show a limited improvement, and one with a burst of errors during the fading period, in which the adaptability of error control is essential. Emerging mobile network quality of service will have a large variation of bandwidth and delay. Thus, adaptive rate control and error recovery are identified as more crucial issues for future research.

Keywords**—***Error control, image compression, mobile networks, rate control, wireless video delivery.*

I. INTRODUCTION

The rapid growth of wireless communication and networking protocols, such as 802.11 and cellular mobile networks, is bringing video into our lives anytime, anywhere, on any device. However, wireless video delivery faces several challenges, such as high error rate, bandwidth variation and limitation, battery power limitation, and so on. This paper presents the latest technology for wireless video content delivery.

Media coding over wireless networks is governed by two dominant rules. One is the well-known Moore's Law, which states that computing power doubles every 18 months. Moore's Law has been at work for codec evolution, and there have been huge advances in technology in the ten years since the adoption of MPEG-2. The second governing principle is the huge bandwidth gap (one or two orders of magnitude) between wireless and wired networks. This bandwidth gap demands that coding technologies achieve efficient compact representation of media data over wireless networks. Obviously, the most essential requirement for "wireless video" is coding efficiency.

Unfortunately, Moore's Law continues to increase transistor budgets. We all know that clock speed is a key means of increasing CPU performance, but every time the clock speed goes up, so does the power. If media coding power dissipation increases beyond a modest 100 mW (that corresponds to Wireless Local Area Network (WLAN) IEEE 802.11 receiver power consumption), it will be hard to implement the media application in portable devices. Hardware solutions are reported in [1], but power control optimization issues remain.

In addition to coding efficiency and power dissipation, error resilience is an important issue, since mobile networks generally cannot guarantee error-free communication during fading periods. Fig. 1 shows a functional block diagram for wireless video communication with respect to error control. The diagram shows four layers: the layer-1/2 transport, which corresponds to physical and media access control (MAC) layers; an end-to-end transport layer such as TCP/IP or RTP/UDP/IP; an error-resilience tool and network adaptation layer such as H.264 NAL [2]; and a source coder layer. To provide error robustness, we can take a number of approaches, combining the four layers. Fig. 2 introduced in [3] illustrates qualitatively the reconstructed video quality depends on redundancy in the data. This is required to protect and recover the transmitted video at the four layers. The figure shows that as the channel error rates increases, a higher amount of redundancy should be allocated for error control. Intuitively, error control in wireless video transmission is concerned with how and what redundancy is allocated to each layer.

The literature on this subject is plentiful. Recent reviews that have appeared in [3], [4], and others cover recent coding technology [5], transport, and error control technologies [6]–[8] in their proceedings. In this paper, therefore, we try to provide a practical insight into wireless video delivery through the investigation of characteristics of existing mobile networks, commercialized or standardized transport, and coding technologies.

Manuscript received January 15, 2004; revised May 31, 2004.
M. Etoh is with DoCoMo Communication Laboratories, San Jose, CA 95110 USA (e-mail: etoh@ieee.org).
T. Yoshimura is with Multimedia Laboratories, NTT DoCoMo, Yokosuka 239-8536, Japan (e-mail: yoshimura@mml.yrp.nttdocomo.co.jp).
Digital Object Identifier 10.1109/JPROC.2004.839605

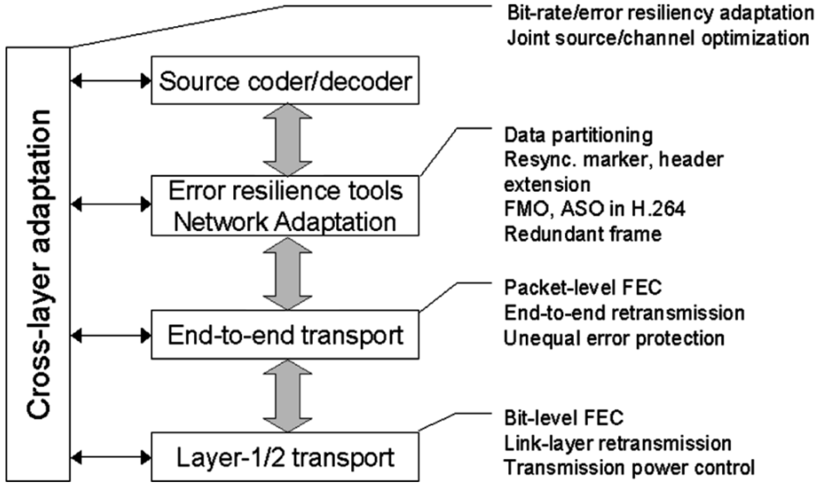

Fig. 1. Functional block diagram for wireless video communication.

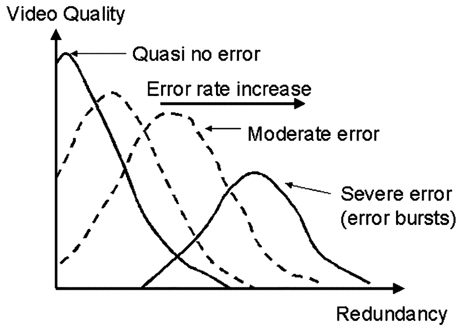

Fig. 2. Video quality versus redundancy in error-prone environments (modified from [3]).

In the following sections, we will examine the redundancy that is added to each layer. This paper is structured as follows. Section II introduces generation of mobile networks and transmission characteristics of recent wireless link technologies. We see that layer-1 and -2 transport tends to provide two distinct conditions: quasi-error-free and burst errors during a fading period (see also "quasi no error" and "severe error" in Fig. 2). Section III provides an overview of end-to-end transport technologies by taking visual telephony, packet streaming services, and multimedia messaging service (MMS) in the Third Generation Partnership Project (3GPP) standards as examples. Section IV introduces MPEG-4 error-resilience tools for bit errors and H.264 network adaptation tools for packet-loss errors. We also evaluate these technologies in view of existing layer-1 and -2 characteristics. Section V discuss future directions for wireless video delivery.

II. Recent Mobile Networks

We have experienced three mobile network generations: analog [first generation (1G)], digital [second generation (2G)], and third generation (3G), which is characterized by its ability to carry data at much higher rates. The technical distinctions and definitions of 1G–3G mobile networks are more fully developed in [9] and [10]. In this section, we see the following observations about the recent and future mobile networks.

- Radio access networks (RANs) using different technologies are going to be integrated into one IP network in the future beyond 3G. The heterogeneity of RANs is the key issue to tackle.
- The state-of-the-art mobile networks try to preserve a very low error rate by controlling available bandwidth. When being out of control limit, they result in periods of high error rate (that is, burst errors).
- According to 3GPP wide-band code division multiple access (W-CDMA) standards, link error quality in toll-quality services is generally very high.

Regarding the heterogeneity of RANs, it is commonly believed that the next generation mobile network [fourth generation (4G)] will operate on Internet technology combined with various access technologies such as WLAN. It will run at speeds ranging from 100 Mb/s in cellular networks to 1 Gb/s in hot-spot networks [11]. In order to ensure connection ubiquity together with high bandwidth and mobility, the network architecture must be heterogeneous rather than homogeneous. 4G technologies include management of handovers (within the same RAN technology, and across different RAN technologies) and, thus, involve alternation of network quality of service (QoS) (e.g., bandwidth, delay). In the following sections, we see how network QoS varies within the same RAN.

A. *QoS Trends: A Large Variation of Bandwidth and Delay*

Table 1 summarizes layer-1 and -2 characteristics of existing RANs. The technical details are beyond the scope of this paper, but note that the trend in the emerging access networks is for use of "adaptive modulation and coding" (AMC).

Fig. 3 shows the principle of AMC. It adaptively changes the level of modulation [e.g., binary phase shift keying (BPSK), quadrature phase shift keying (QPSK), 8-PSK, 16-quadrature amplitude modulation (QAM)] and so on) and amount of redundancy for error correction code. Higher level of modulation (e.g., 16-QAM) with no error correction code can be used by users with good signal quality (close to the base station) to achieve higher bandwidth. Lower level of modulation (e.g., BPSK) with more redundancy for error correction is used by users with bad signal quality (in cell edge) to keep the channel condition, accepting

Table 1
Layer-1/2 Comparison of RANs

Radio Access	PHY	Channel Bandwidth	Modulation	Channel Coding	Error Recovery
W-CDMA [12]	64 - 384 Kbps	5 MHz	QPSK (downlink) BPSK (uplink)	Convolutional or Turbo Code	SR-ARQ
HSDPA [13], [14]	64 Kbps - 14 Mbps (downlink)	5 MHz	QPSK 16QAM (downlink)	Rate 1/3-1 Turbo Code based	Type-I HARQ with chase combining, Type-II HARQ
CDMA-2000 [12]	1.2 - 307.2 Kbps(1X)	1.25 MHz(1X) 5 MHz(3X)	QPSK(downlink) BPSK (uplink)	Convolutional or Turbo Code	SR-ARQ
CDMA-2000 1x EV-DO [15]	Downlink peak rate: 1.25 - 2 Mbps Uplink peak rate: 144 Kbps	1.25 MHz	QPSK, 8PSK, 16QAM (downlink) BPSK (uplink)	Convolutional or Turbo Code	Type-II HARQ
GPRS [16]	9.06 - 171.2 Kbps	200 KHz	GMSK	Convolutional Code for CS1-4 mode none for CS4 mode	SR-ARQ FEC
EDGE [17]	8.8 to 473.6 Kbps	200 KHz	GMSK 8PSK	Convolutional Code (CS1-4, MCS1-9)	HARQ II
801.11 [18], [19], [20]	6-54 Mbps (11a) 1-11 Mbps (11b) 1-54 Mbps (11g)	20-22 MHz	OFDM (11a) CCK (11b) OFDM+PBCC(11g)	Convolutional	SW-ARQ

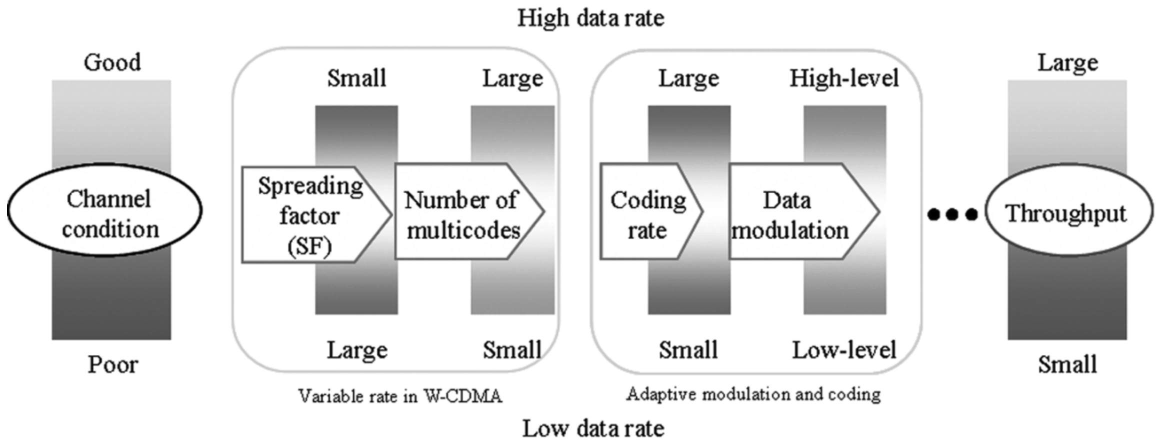

Fig. 3. HSDPA layer-1-2 control.

lower bandwidth available. In high-speed downlink packet access (HSDPA), for example, turbo code is used for error correction code with a rate varying between 1/3 and 1, and QPSK, 8-PSK, 16-QAM, and 64-QAM can be used for its modulation scheme.

In addition to AMC, HSDPA has the following technical features.

- Hybrid automatic repeat request (H-ARQ): This is a combination of retransmission control (ARQ) and forward error correction (FEC). When retransmitting a packet, either Type-I or Type-II H-ARQ can be used. The original packet is retransmitted in Type-I H-ARQ, and a redundant packet that is different from the original packet is transmitted in Type-II H-ARQ.
- MAC layer packet scheduling: Since multiple mobile terminals share the same HSDPA channel, a base station can flexibly schedule which terminal's packet to send, considering fairness and total throughput.

These technologies enable high throughput up to approximately 14 Mb/s, and we expect this to facilitate high bit-rate video service in a wireless environment.

Although this discussion focuses on the control of HSDPA, this concept could be equally applicable to 1xEV-DO and WLAN. The idea is to limit the number of link errors by adjusting the dedicated bandwidth through the AMC. The difference is that W-CDMA also adjusts the spreading factor and the number of multiplexing spreading codes. Regarding adaptive modulation, 1xEV-DO and WLAN have a similar strategy. Typically, the average bit-error rate (BER) requirement is set beforehand, depending on the class of application. The adaptive modulation is applied to ensure the QoS that is evaluated directly by signal-to-interference ratio (SIR) or indirectly measured by BER. Layer-1/2 transport control generally provides two distinct states, quasi-error-free and burst errors during fading periods, while there is a large variation of bandwidth and delay.

B. QoS Control Examples: W-CDMA

The protocol architecture of W-CDMA is shown in Fig. 4, [21]. Only the U-plane, which transmits user data, is shown in this figure; the C-plane, which deals with control signals is omitted for simplicity. The upper part of the figure shows typical application examples in W-CDMA. The next part represents layer-2, composed of three sublayers, packet data convergence protocol (PDCP), radio link control (RLC), and MAC. The lower part shows layer-1 physical channels (PHY).

There are two types of services, circuit switched and packet switched. Voice call, video call using 3G-324M

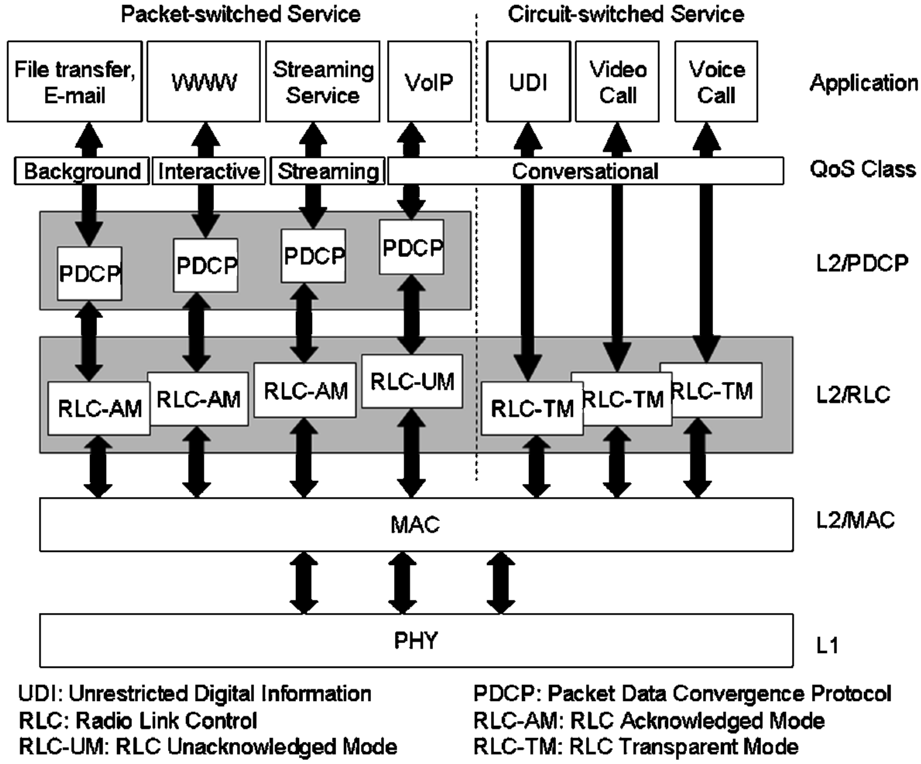

Fig. 4. W-CDMA L1-2 protocol stack.

(described in the next section), and unrestricted digital information (UDI) bearer service (equivalent to ISDN bearer service) are examples of circuit-switched services. Voice over IP (VoIP), multimedia streaming, and data service including messaging and Web browsing are examples of packet-switched services. The PDCP sublayer, which is applied only to packet-switched services, works to hide the differences of higher layer protocols from the lower layers, and provides services such as header compression for efficient packet delivery.

The RLC sublayer provides fragmentation/reassembly mechanisms from the data unit of the upper layers (that is, IP) into RLC-PDU, which is the data unit of W-CDMA transmission. It also provides error detection, orderly packet delivery, link-layer retransmission, and so on. It offers the following three transmission modes.

- Transparent mode: This mode provides only fragmentation/reassembly mechanisms without any header information and padding. This is mainly used for the circuit-switched service.
- Unacknowledged mode: This mode performs error detection as well as fragmentation/reassembly. The data units in which bit errors are detected are dropped at the receiver side. Since it offers no link-layer retransmission, it can be applied only to a packet-switched service that has strict delay restriction.
- Acknowledged mode: In addition to fragmentation/reassembly and error detection, this mode provides link-layer retransmission and orderly packet delivery. A packet-switched service that allows some delay can use this mode.

The MAC sublayer provides each RLC entity with a logical channel, and maps the logical channels to layer-1 physical channels. The MAC sublayer also performs transmission priority control according to the RLC buffer level and monitors the amount of traffic.

Layer-1 provides various physical channels based on service requirements and/or the radio resource condition. The physical channels include the dedicated channel occupied by one mobile terminal, the common channel shared by multiple terminals, the HSDPA channel, and so on.

In order to support various types of services, 3GPP defines four QoS classes; conversational, streaming, interactive, and background [22]. The QoS class used for each service type is determined by network operators. Typically, delay-sensitive services including circuit-switched service and VoIP are assigned to the conversational class, packet-streaming services requiring a certain bit-rate guarantee are assigned to the streaming class, and best-effort and delay-insensitive services such as Web browsing and messaging are in the interactive class or the background class. QoS class assignment is initiated by the terminal. The terminal requests the network layer to set up network layer bearers. In GPRS and W-CDMA packet networks, this process is called Packet Data Protocol (PDP) context activation. According to the 3GPP QoS model, different terminal applications may request differentiated QoS from the network according to the application needs, assuming that the network supports the QoS class.

As for the selection of unacknowledged and acknowledged modes, Fig. 4 shows just one example, mapping of services to the RLC modes through PDCP, which is an operator's implementation matter. Generally speaking, layer-2 retrans-

Table 2
QoS Suggested by 3GPP Standards

Service	Bit Error Rate	Transport Packet Loss Rate
UDI bearer	1e-6	NA
AV bearer	1e-4 or better	NA
Packet-Switched conversation	NA	1e-3
Packet-Switched streaming	NA	1e-4 or better

mission works only for wireless link errors, while application layer-7 retransmission works for both packet loss due to network congestion and wireless link errors. If a moderate amount of delay, say one second, is allowed, the acknowledged mode should be used, since a typical round-trip time (RTT) at W-CDMA layer-2 is typically below 100 ms while that of layer-7 RTT is several hundred ms. If the service allows a longer delay, layer-7 retransmission should also be considered. We will see its usage later as RTP retransmission.

Among the above-mentioned services, the video-related services require higher channel quality compared with the values seen in a number of research papers. For example, according to the results in [23], the quality of a video call using 3G-324M in circuit-switched service (AV bearer) shows unacceptable scores in the neighborhood of BER 1e-3, and a BER 1e-4 or better channel quality is required. According to [24], a packet-loss rate (PLR) of 1e-3 at transport layer is required for packet-switched conversational service (PSC). For these services, the required quality is satisfied by the transmission power control (TPC). Packet-switched streaming service (PSS), on the other hand, requires PLR 1e-4 or below [25] at transport layer. Since the streaming service tolerates a delay of up to a few seconds, the required PLR is easily satisfied by the link-layer retransmission of RLC acknowledged mode. Moreover, because of compatibility with ISDN bearer service, a UDI bearer service should satisfy BER 1e-6. Table 2 summarizes the required QoS in terms of link error.

We believe that the channel quality described above is necessary for providing toll-quality video service (in other words, those show practically operational QoS) and that video transport technologies which show a limited improvement only in poor radio conditions (more than a few percent of PLR), do not make any industrial sense. Currently, the best way to provide toll-quality video service is to allocate the redundancy mainly to layer-1/2 and secondarily to end-to-end transport such as H.223 and Real-Time Transport Protocol (RTP)/RTP Control Protocol (RTCP). If we develop new upper layer technologies that guarantee toll-quality service, even in the environment of a few percent PLR, the redundancy should vary so that it can lead to a more efficient radio resource utilization. This discussion will be also related to cross-layer optimization.

III. Existing Transport Standards for Wireless Video Delivery

This section introduces examples of end-to-end transport layer technologies, illustrated in Fig. 1. 3GPP has standardized three types of visual content delivery services and technologies. These are:

1) circuit-switched multimedia telephony; [26]
2) transparent end-to-end packet-switched streaming [27];
3) MMS [28].

A. Multimedia Telephony Service

The 3G-324M standard was developed for circuit-switched multimedia telephony service (visual phone), and it is applied to speech [adaptive multirate (AMR)] and video codecs, the communication control unit, and the multimedia multiplexing unit. The video codec requires the H.263 baseline as an essential capability and recommends MPEG-4 support. The support of H.223 Annex B, which offers improved error resilience, is mandatory for the multimedia multiplexing unit. While various media coding schemes can be used in 3G-324M by exchanging the terminal capability through the use of communication control procedures, and by changing the codec setting upon the establishment of logical channels, 3G-324M defines a minimum essential codec to ensure interconnection between different terminals. For the video codec, 3G-324M specifies the H.263 baseline (excluding the optional capabilities) as the essential video codec (as is the case for H.324). It also recommends the use of MPEG-4 video to deal with transmission line errors unique to mobile communications. For error resilience, multimedia multiplexing functionality is enhanced in this standard. The multiplexer also provides transmission service according to the type of information (such as QoS and framing). H.223, the multimedia multiplexing scheme for H.324, satisfies the requirements by adopting a two-layer structure consisting of an adaptation layer and a multiplexing layer. The 3GPP extension adds error-resilience tools step by step to ITU-T's original H.223, so that error-resilience levels can be selected according to the transmission line characteristics.

Three adaptation layers are defined according to the type of the higher layers.

- AL1: For user data and control information. Error control is performed in the higher layer. Iterative retransmission will be applied for reliable transmission.
- AL2: For voice. Error detection and sequence numbers can be added.
- AL3: For video. Error detection and sequence numbers can be added. ARQ is applicable.

NTT DoCoMo, Tokyo, Japan, launched visual phone service in 2001 for their 3G network, W-CDMA, where it provides a 64-kb/s circuit switch connection that is compatible with N-ISDN. Note that H.223 contains fundamental and essential error control features: synchronization, error detection, FEC, ARQ, and data duplication. These are combined for better error resilience depending on the type of payload: data, voice, or video. H.223 is designed mainly for an error prone circuit switch connection environment, although its design is general in terms of payload adaptation and associated error control. For further technical discussion see [29].

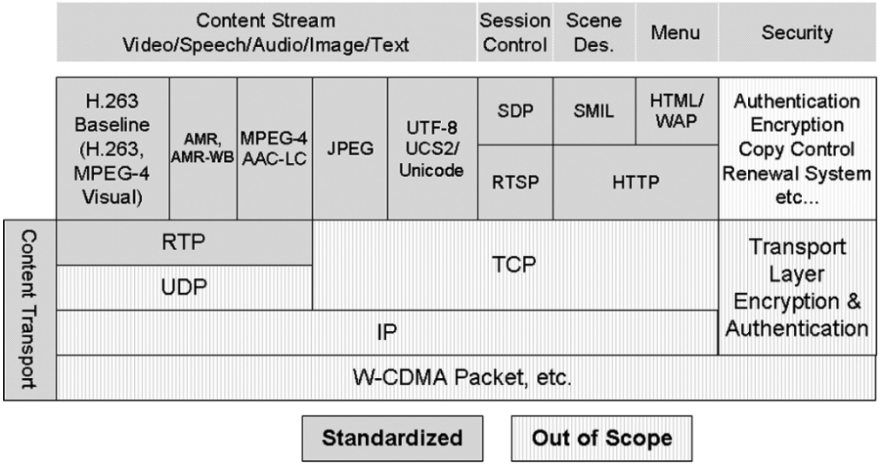

Fig. 5. 3GPP packet streaming.

B. PSS

When considering most typical applications over the Internet and IP-transparent seamless connectivity, media streaming will certainly be based on a packet switch connection, whereas a combination of Internet Engineering Task Force (IETF) standards such as RTP [30] will be used over mobile networks. Packet-based streaming services are required whenever instant access to multimedia information must be integrated into an interactive media application, so that a Web server can work with requests for information, deliver that particular information as fast as possible, complete the transaction, disconnect, and go on to other requests. A client connects to a Web server only when it needs information. A 3GPP standard, transparent, end-to-end PSS (3GPP PSS) has been specified to fulfill these considerations.

Fig. 5 shows packet-based streaming protocol stacks standardized in 3GPP for a W-CDMA network. The media types specified in this standard are very similar to 3G-324M. For session control, however, RTSP [31] and SDP [32] are used. For the media transport protocol, RTP is used. RTCP reports to the sender the reception status of images and audio transmitted with RTP to control the service quality.

Media transport technologies should differ for each type of wireless link. For real-time video conferencing over the circuit-switched network, we use 3G-324M for the error control. H.223 is a very sophisticated multiplexing scheme for bit errors, but it cannot be used for packet streaming in which packet loss (packet erasure) is a problem. For robustness with respect to packet loss, we must abandon the idea of "joint source-channel decoding" (the H.324 approach), where the receiver tackles bit-erroneous packets using source coding knowledge.[1] Moreover, streaming services are allowed to have somewhat more latency, while real-time video conferencing has a strict latency constraint, such as 200 ms. Among the error control fundamentals, synchronization and error detection are solved by an underlying transport-layer protocol stack such as RTP/UDP. FEC, ARQ, data duplication and other techniques such as data interleaving and unequal error-resilience packetization are relevant in some ways to packet-loss resilience [35]. We will discuss those techniques later, and focus on the standardized protocols here to see how the error control protocols work for packet-based media streaming.

Real-time media streams that use RTP are, to some degree, resilient against packet losses. RTP provides all the necessary mechanisms to restore the ordering and timing that are present at the sender in order to properly reproduce a media stream at the recipient. RTP also provides continuous feedback about the overall reception quality from all receivers, thereby allowing a sender in the midterm (on the order of several seconds to minutes) to adapt its coding scheme and transmission behavior to the observed network QoS. To give a packet-based streaming server finer error control, an extension of RTCP [36], [37] has been standardized and is now being adapted to the 3GPP PSS. With this extension, RTP makes a provision for timely feedback that would allow a sender to repair the media stream immediately through retransmissions, retroactive FEC control, or media-specific mechanisms for some video codecs, such as reference picture selection. For example, we can realize a retransmission method [38] for RTP applicable to unicast and (small) multicast groups.

All necessary protocols underlying media adaptation are included in 3GPP PSS and the emerging IETF standards. Retransmission control and media-specific mechanisms are left intact as not standardized. We will examine specific mechanisms in the following sections.

C. MMS

In MMS, multimedia content is delivered to the user asynchronously by means of a message. At the time of writing (late 2004), mobile picture mail services including MMS have proved a major hit in recent mobile markets. The picture mail service allows users to transmit still images from compatible mobile phones with built-in digital cameras to virtually any device capable of receiving e-mail. As one of the 3G services, the mail service is extended in Japan to enable

[1]The exception is the approach taken by proponents of the UDP-Lite Protocol [33] and related schemes [34]. They proposed UDP-Lite as benefiting from having damaged data delivered rather than discarded by the network, where an intermediate layer prevents error-tolerant applications from running well in the presence of bit errors. Roughly speaking, UDP-Lite is equivalent to AL2 and AL3 of H.223.

users to e-mail video clips as large as several hundred kilobytes taken either with the mobile device's camera or downloaded from sites. For this application, ISO/IEC and 3GPP standards have been adopted. The media types specified in MMS are text, AMR for speech, MPEG-4 AAC for audio, JPEG for still images, GIF and PNG for bitmap graphics, H.263 and MPEG-4 for video, and scalable vector graphics (SVG). The major technical issues in MMS are compression efficiency and multimedia content wrapper (file format). The file format specified by the MMS standard is the 3GGP file format, which has the 3GPP file extension and is based on the ISO media file format. A 3GPP file can contain multiple media types but supports only one video track, one audio track, and one text track for simplicity of implementation.

Apart from asynchronous messaging, there are two methods of distribution between the multimedia information distribution server and a mobile device: the streaming method and the download method. The 3GPP standard-based file format is now being also used for content download service. The download method can use the file format for the distribution of media clips. To minimize an initial playback latency, the file format also supports progressive download, which allows a terminal to start playback before the file is fully downloaded. To do this, the media tracks must be interleaved properly within the file, so that the client receives short portions of each media type in turn. Here, the download method requires a reliable communication protocol between the multimedia information distribution server and a terminal, even though some transmission delay may be tolerable. Communication procedures that meet this requirement include HTTP on TCP/IP, which is used widely over the Internet.

There is a very interesting issue in TCP/IP file transfer over wireless links. A typical RTT in recent mobile networks varies between a few hundred milliseconds and 1 s, due to effects from other users and from mobility [39]. Arriving and departing users can reduce or increase the available bandwidth in a cell. Increasing the distance from the base station decreases the link bandwidth due to reduced link quality, and by simply moving into another cell, the user can experience a sudden change in available bandwidth. Changing from a fast to a slow cell is normally handled well by TCP's self-clocking property. However, if upon changing cells, for example, a connection experiences a sudden increase in available bandwidth, that bandwidth can be underutilized, because TCP increases the sending rate slowly for congestion avoidance. A sudden increase in RTT can cause a spurious TCP timeout. A delay spike is a sudden increase in the latency of the communication path. Cellular links are likely to experience delay spikes exceeding the typical RTT by several times due to the link layer recovery, handover, and blocking by high-priority traffic. In addition, a large TCP window used in the fast cell can create congestion resulting in overbuffering in the slow cell. In order to avoid a lengthy recovery through retransmission timeouts, it is necessary to specifically tune for a window large enough for efficient loss recovery. IETF has specified a management scheme as a best current practice. For the details, see [39].

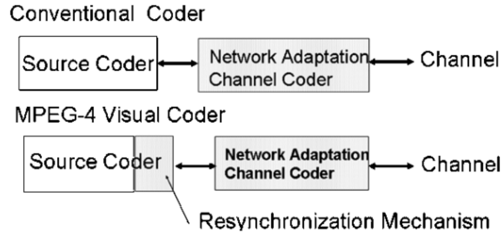

Fig. 6. MPEG-4 error-resilience structure.

IV. Error-Resilience Tools in Existing Video Standards

In the previous sections, we discussed standardized video content delivery services over mobile networks, focusing on transport and layer-1/2 technologies. Now we will discuss error control and recovery technologies currently in use or planned in the near future at source code layer.

A. MPEG-4 Video

MPEG-4 video[2] error-resilience tools were developed on the assumption that there is a benefit in having damaged data delivered from H.223. At the source coder layer, these tools provide synchronization and error recovery functionalities. Fig. 6 summarizes MPEG-4 video.

- Resynchronization maker: This localizes transmission errors by inserting resynchronization code in variable-length coded data at the source coder layer.
- Data partitioning: This enables error concealment by inserting synchronization code at boundaries where different types of data (e.g., motion vectors and discrete cosine transform (DCT) coefficients) meet in coded data.
- Reversible Variable Length Code: Variable length code that can be decoded from the reverse direction is applied to the DCT coefficient.
- Adaptive intrarefresh (AIR): This tool prevents error propagation by performing intraframe coding frequently on motion domains.

When considering the BER shown in Table 2 and burst bit errors owing to W-CDMA turbo code, practical and efficient tools are restricted. Recall the discussion on the two distinct error states. Although AIR is not a normative part of MPEG-4 standard, it works well, especially when burst error occurs. Another beneficial tool is the resynchronization marker because of its simplicity and adaptability to burst errors. For further details, see [40] and [41].

B. H.264

There is considerable literature on the error resiliency future of H.264/AVC (in short, H.264)[3] [2], [42]. H.264 is based on hybrid video coding, and is similar in spirit to other standards such as MPEG-4, but with new coding technologies:

[2]The official name is ISO/IEC 14 496-2.

[3]ITU-T H.264 is also known as MPEG-4 Part 10, MPEG-4 advanced video coding (AVC), and ISO/IEC 14 496-10 in official MPEG terminology. For simplicity, we refer to H.264.

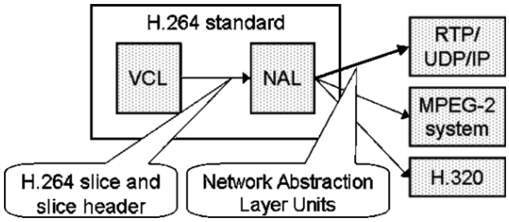

Fig. 7. NAL.

multimode and multireference MC, fine motion vector precision, B-frame prediction weighting, 4×4 integer transform, multimode, intraprediction, in-loop deblocking filter, uniform variable length coding (UVLC), network abstraction layer (NAL), switching picture slices (SP slices), and so on. Because of these new coding technologies, H.264 has the highest performance at the time of this writing (late 2004). It is said that H.264 attains substantial bit-rate savings (up to 50%) relative to other standards such as MPEG-4 at the same visual subjective quality. The decoding complexity at least doubles that of MPEG-4, while the encoding complexity may quintuple that of MPEG-4 or more, depending on the degree of rate-distortion (R-D) optimization and complexity of motion estimation. It may cause power dissipation in case of software encoding. Since other papers cover H.264 in detail, let us focus on the error-resilience tools of H.264. H.264 consists of two layers.

- Video coding layer (VCL): contains the signal processing functionality (transform, quantization, motion search/compensation, loop filter, and so on.)
- NAL: encapsulates the slice output of the VCL into NAL units suitable for transmission over packet networks.

In contrast with the MPEG-4 error-resilience structure shown in Fig. 6, the H.264 error-resilience structure is based on a flexible network adaptation structure called NAL,[4] shown in Fig. 7. Unlike MPEG-4, NAL elegantly separates the H.264 source coder tasks from error-resilience tasks, such as data partitioning. Moreover, the H.264 error-resilience structure is based on the assumption that bit-erroneous packets have been discarded by the receiver. The error-resilience structure has been designed mainly for packet-loss environments, and this design concept is valid, since Internet protocols are currently widely used for content delivery. According to [42], the following features of NAL create robustness with respect to data errors and losses, and flexibility of operation over a variety of network environments.

- Parameter set structure: Parameter sets are exchanged reliably by out-of-band signaling [for example, Session Description Protocol (SDP)] in advance. An in-band method is also available, but is not intended. A parameter set includes picture size, display window, macroblock allocation map, and so on.
- Flexible macroblock ordering (FMO): This allows transmission of a macroblock in nonraster scan order so that we can apply spatial concealment.
- Arbitrary slice ordering (ASO): This capability can improve end-to-end delay in real-time applications, particularly when used on networks having out-of-order delivery behavior (e.g., the Internet).
- Redundant pictures: Redundant slices and spare macroblocks provide additional protection against the propagation of errors via interprediction.
- Data partitioning: This design shares the MPEG-4 data partitioning concept. It may produce better visual quality if the underlying network provides unequal error protection based on the priority of the packet.
- SP/SI synchronization/switching pictures: This is a feature unique to H.264 that allows switching a decoder between representations of the video content that use different data rates. It contributes to error recovery as well as trick modes.

The introduction of NAL is a major advance from the MPEG-4 error-resilience structure. The NAL unit syntax structure allows greater customization of the method of carrying the video content to the transport layer. A typical example is packetization to RTP payload [43]. For example, the flexible syntax with ASO enables robust packet scheduling, where we send important packets earlier and retransmit lost packets so that we can improve the transmission reliability for important pictures. The H.264 NAL structure together with underlying transport technologies such as RTP can support most of the packet-loss resilience schemes summarized in the next section.

Of the error-resilience tools newly introduced in H.264, slice interleaving enabled by FMO and ASO is one of key features. However, there is some concern that handling of this FMO may complicate the implementation of decoder buffer management significantly, while providing benefits only in the case that slices are randomly and moderately lost. Switching between FMO and non-FMO coding may be desired depending on the link condition, if the complexity allows.

V. Discussion: Future Directions for Wireless Video Delivery

Earlier in this paper, we reviewed the four layers of wireless video delivery technologies: the layer-1/2 transport in Section II, the end-to-end transport layer in Section III, the error-resilience tools and network adaptation layer, and the source coder in Sections III and IV, respectively. The description of error control technologies is a large part of this paper. Without limiting or restricting the discussion to error control, there are a number of related promising technologies for wireless video delivery.

A. Summary of Related Technologies

To sum up our practical review of recent wireless video delivery technologies, Table 3 provides a set of possible delivery technologies.[5] Note that the end-to-end transport technologies, such as H.223, include synchronization and error

[4]Network adaptation is termed "network abstraction" in H.264.

[5]Cross-layer technologies must exist aside.

Table 3
Possible Wireless Video Delivery Technologies

layer	video telephony	packet streaming	messaging + progressive download
source coder	error concealment feedback-based error control (Adaptive Intra Refresh) Reversible VLC	error concealment feedback-based error control	
error-resilience tools network adaptation	resync marker data partitioning selective ARQ(AL3)	slice interleaving data partitioning redundant pictures packet scheduling selective ARQ	interleaving for progressive download
end-to-end transport	H.223	RTP/UDP+RTCP +Header Compression [44]	reliable transport such as wireless-TCP/IP [39]
Layer-1/2 transport	FEC (BER:1e-4)	FEC+ARQ (PLR:1e-4)	FEC+ARQ (PLR: $<$ 1e-4)

detection, and can provide FEC and ARQ, if necessary. As for the source coders, we assume that MPEG-4 will be selected for video telephony and H.264 will be used for packet streaming.

Error concealment has a long history; it has been available from H.261 and MPEG-2. The easiest and most practical approach is to hold the last frame that was successfully decoded. The best-known approach is to use motion vectors that can adjust the image more naturally when holding the previous frame [45]. The vector approach works well at a relatively high bit rate, because the amount of motion vector data is small, and moreover, the motion vectors represent realistic object motions on the image. More sophisticated error concealment methods consist of a combination of spatial/spectral and temporal interpolations with motion vector estimation [35]. Object-based error concealment is found in [46]. This research uses *a priori* information for concealment through a principal component analysis on models.

One can also add optimization techniques to **feedback-based error control**, especially for streaming services in which a relatively large RTT is allowed. As for the feedback-based error control, Girod and Faerber reviewed those technology in depth [47]. According to their review, feedback-based error control is basically for error mitigation by feedback. A feedback channel indicates which parts of the bit stream were received intact, and which parts of the video signal could not be decoded and had to be concealed. Having this feedback, the source coder typically uses the INTRA mode for some macroblocks to stop interframe error propagation. Originally, selective recovery of video packets [48] in AIR works well. Reference picture selection is another well-known feedback-based control technique performed at the source coder layer. With data partitioning, we assume two conditions.

- The source bit stream can be divided and prioritized in terms of significance to the reconstructed image quality.
- The underlying transport provides **unequal protection** by differentiating FEC, ARQ, data duplication, transmission sequence, and so on.

We cannot use the data partitioning approach if all the portions of a bit stream are equally important for image reconstruction.[6] When the two conditions hold, we can apply data partitioning. Many error control schemes are categorized into this "unequally partitioned bitstream—unequal protection mapping" framework. MPEG-4's data partitioning of motion vectors and DCT coefficients, prioritized I-picture, and slice transport are typical examples [49]. In a wider sense, H.264 parameter sets (such as picture size and coding mode), which exchange using reliably by out-of-band signaling, can fall into this category. The **layered or scalable coding** approach [50]–[52] is another refinement of data partitioning. These approaches might be used in a narrow-band wireless broadcasting scenario.

Cross-layer optimization, which is not listed in the summary table, has recently become of interest. One such approach is called **joint source-channel coding** [53], in which two distortion measures (quantization distortion introduced in source coding and channel distortion caused by channel errors) are formalized and modeled in associated with coding rate, channel errors, and concealment redundancy (in this case, intra refreshing rate). An R-D optimization scheme is applied over these measures.

Packet scheduling, which uses source information dependencies [54], is another cross-layer optimization. For this type of optimization, interframe prediction dependency at source coder layer is mapped to packet dependency at the end-to-end transport layer. Given constraints such as jitter and a packet-loss model or feedback information, packet processing is optimized. Optimization can be realized in two ways: client-side optimization or server-side optimization. Adaptive media playout [55] is the client-side process and can be categorized as another type of error concealment. This technique allows a streaming media client to control the rate at which data is consumed by the playout process. When a frame is lost due to the packet loss, the decoder can slow the playout to reduce the visual impairments, on which R-D optimization based on packet dependency is applied.

As for the server-side optimization, R-D optimized packet transmission scheduling, which was recently proposed in [56], has gained the position of an established packet streaming framework. The framework consists of mapping the source coder information (e.g., interframe dependency) to underlying packet structure and then scheduling the

[6]A layered or scalable coding approach is a convenient way to create this condition.

packet transmission based on the packet dependency with given constraints. Its practical usage was also confirmed in [57].

If we extend the joint source-channel coding framework to include W-CDMA layer-1/2 transport, we can manage **transmission power optimization** together with R-D optimization. These contributions are found in [58]–[61]. In contrast with error control, the number of papers that consider power dissipation is small. Recent work [58] involves processing power for source coding and channel coding as well as the transmission power, jointly. According to the current mobile network operations and implementations, the layer-1/2 transport control is isolated from upper layers, except through the predefined QoS class designation, such as the PDP context activation. Future networks might provide a programmable function for layer-1/2. In that case, a key issue will be cross-layer optimization, including layer-1/2 associated power consumption. In addition, if necessary, other key parameters will include spreading factor, number of multiplexing spreading codes, coding rate of error correction code, and level of modulation.

B. Future Directions

In Section II, we observed that RANs are going to be heterogeneous. In addition we noted that current RAN implementations offer a very low error rate QoS (approximately 1e-4 PLR), because the layer-1/2 error control keeps error QoS as low as possible.

We believe that emerging mobile networks will have the following characteristics.

- Large bandwidth and delay variation: layer-1/2 transport results in large variations of RTT, jitter, and bandwidth, while trying to keep error rate constant. There are some exceptions such as W-CDMA circuit switch connections, if a constant bandwidth with low latency is required.
- Two distinct error states: Error conditions are bisected into a good stable state and burst error state.

From these characteristics, we can conclude that the essential error control features are:

- extraordinary adaptability to burst error state;
- rate control in the two-state error prone environment, as identified above

Referring back to Fig. 2, the issues are how to efficiently detect the state transition from "quasi no error" to "severe error" and how to adopt the delivery system to the state change over heterogeneous RANs. For error recovery and concealment in the burst error state, the essential element is rate adaptation to introduce data redundancy.

In our opinion, many researches focus on intermediate error conditions, which results in moderate error control schemes highly optimized to hypothesized error patterns. This is not sufficient. Extraordinary adaptability definitely requires a combination of capabilities, such as *feedback-based control* and *error concealment*. For such adaptive error recovery, as we saw in packet transmission scheduling, it will be more promising to control the error at the transport layer or below, where the source coder information is mapped to the underlying packet control. This is because of system scalability.

There are many research contributions concerning rate control over mobile networks [62]–[64]. These rate-control technologies need to be evaluated with respect to the two-state error prone environment. Concerning rate control over mobile networks, we already discussed that rate-control mechanisms utilize the RTCP. All of the rate-control mechanisms mentioned above assume that packet loss, delay, and jitter are caused by network congestion. In mobile networks, however, packet loss and jitter may also be caused by radio link errors. In addition to the higher BER, cellular links cause large transmission delays, sometimes more than 100 ms. When conventional rate-control mechanisms are applied to mobile networks, a sender cannot identify the network congestion condition correctly, and this leads to inappropriate rate control. A typical symptom is that a sender reduces its transmission rate even if the network is not congested. In TCP, several techniques have been proposed to counter this issue [65], including local retransmission [65], split-connection [66], and explicit loss notification (ELN) [67]. Some of them can be also applied to RTCP. For example, a split-connection approach to RTP is proposed in [68]. That puts a middle person at midway between wired networks and radio links so that the servers can realize both wired and wireless network QoS separately. ELN is a cross-layer approach by signaling link layer information to transport layer. A more generic cross-layer framework is discussed in [69], where such signaling can contribute to more efficient congestion control and QoS adaptation by avoiding duplicated effort across layers. Although a framework has been discussed, cross-layer video rate control such as layer-2/3 interaction with RTP retransmission and layer-2 ARQ has not been reported, as far as we know. We can expect further contributions in this area.[7] Note that a *cross-layer approach* can contribute to the above mentioned adaptability over wireless link as well as power dissipation optimization. Interlayer interfaces like H.264 NAL can be extended in that context.

VI. Conclusion

We reviewed practical video delivery technologies through the examination of existing mobile networks, commercialized or standardized transport, and coding technologies. Compression efficiency, power dissipation, and error control are identified as intrinsic issues in wireless video delivery. Among these issues, we evaluated the four layers of error control technologies. We pointed out that layer-1/2 transport tends to provide two distinct conditions: quasi-error-free and burst errors during fading periods. In the former condition, upper layer error control technologies have a limited role. When considering these role, extraordinary adaptability of error control to the latter condition is essential. We also saw that the emerging mobile network QoS is going to have a larger variation of bandwidth and delay. Thus, we identified

[7]Another design goal is TCP-friendliness [70]. However, it is beyond the scope of this paper.

rate control as the essential technology that provides extraordinary adaptability to varying bandwidth. Cross-layer optimization and rapid feedback-based source coder control with error concealment, specifically for that rate control, are frontiers to be investigated. In addition to those technologies, the heterogeneity of RANs will require the investigation of multipath media delivery technologies, where servers and clients are connected over different networks simultaneously with layer-2 multipaths.

Future review includes multimedia broadcast/multimedia service (MBMS), which is not covered in this paper. MBMS is designed to offer an efficient way to transmit data from a single source to multiple destinations over a radio network. For broadcast and multicast, we may need different techniques that were not addressed here.

REFERENCES

[1] P.-C. Tseng, Y.-C. Chang, Y.-W. Huang, H.-C. Fang, C.-T. Huang, and L.-G. Chen, "Advances in hardware architectures for image and video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 184–197, Jan. 2004.

[2] T. Stockhammer, M. M. Hannuksela, and T. Wiegand, "H.264/AVC in wireless environments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 657–673, July 2003.

[3] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

[4] *IEEE Commun. Mag. (Special Section On Wireless Video)*, vol. 36, no. 4, pp. 92–151, Apr. 1998.

[5] G. Sullivan and T. Wiegand, "Video compression—From concepts to the H.264/AVC standard," *Proc. IEEE*, vol. 93, no. 1, pp. 18–31, Jan. 2004.

[6] Y. Wang, A. R.. Reibman, and S.. Lin, "Multiple description coding for video delivery," *Proc. IEEE*, vol. 93, no. 1, pp. 57–70, Jan. 2004.

[7] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 71–83, Jan. 2004.

[8] Q. Zhang, W. Zhu, and Y.-Q. Zhang, "End-to-end QoS for video delivery over wireless internet," *Proc. IEEE*, vol. 93, no. 1, pp. 123–134, Jan. 2004.

[9] H. Kaaranen, A. Ahitainen, L. Laitinen, S. Naghian, and V. Niemi, *UMTS Networks: Architecture, Mobility, and Services*. New York: Wiley, 2001.

[10] K. Tacikawa, Ed., *W-CDMA: Mobile Communications System*. New York: Wiley, 2002.

[11] "Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000," Int. Telecommun. Union-Radiocommun. (ITU-R) Working Party 8F, Geneva, Switzerland, Recommendation M.1645, 2003.

[12] M. Sawahashi, K. Higuchi, H. Ando, and Y. Okumura, "W-CDMA technology," *NTT DoCoMo Tech. J.*, pp. 60–75, July 2001.

[13] (2002, Mar.) High speed downlink packet access: Physical layer aspects (release 5, TR 25.858 V5.0.0). Technical Specification Group Services and System Aspects, 3rd Generation Partnership Project (3GPP), Sophia-Antipolis, France. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/25_series/25.858/25 858-500.zip

[14] A. Das, F. Khan, A. Sampath, and H.-J. Su, "Design and performance of down link shared control channel for HSDPA," in *Proc. Int. Symp. Personal, Indoor, and Mobile Radio Communications*, vol. 3, 2002, pp. 1088–1091.

[15] H. Okinaka, "3GPP2-the partnership project for the global cdma2000 specifications," in *2003 CDMA Americas Congr.*, Dec. 2003.

[16] "Digital cellular telecommunications system (phase2+); General Packet Radio Service (GPRS); service description; stage 2 (GSM03.60 version 7.4.1 release 1998)," Eur. Telecommun. Standards Inst. (ETSI), Sophia-Antipolis, France, EN 301 344 V.7.4.1, Aug. 2003.

[17] "Digital cellular telecommunications system (phase 2+); GSM/EDGE radio access network (GERAN) overall description; stage 2 (3GPP TS43.051 version 5.10.0 release 5)," Eur. Telecommun. Standards Inst. (ETSI), Sophia-Antipolis, France, TS 143 051 V5.10.0, Sep. 2000.

[18] "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," Int. Standards Org./Int. Electrotech. Comm. (ISO/IEC), ISO/IEC 8802-11, 1999.

[19] *Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: Higher-speed physical layer extension in the 2.4 GHz band*, IEEE Standard 802.11b-1999.

[20] *Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: Higher-speed physical layer extension in the 5 GHz band*, IEEE Standard 802.11a-1999.

[21] K. Tachikawa, Ed., *W-CDMA Mobile Communications System*. New York: Wiley, 2003.

[22] (2003, Dec.) Quality of service (QoS) concept and architecture (TS 23.107). Technical Specification Group Services and System Aspects, 3rd Generation Partnership Project (3GPP), Sophia-Antipolis, France. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/23_series/23.107/

[23] (2001, Mar.) QoS for speech and multimedia codec; quantitative performance evaluation of H.324 annex c over 3G (TR 26.912). Technical Specification Group Services and System Aspects, 3rd Generation Partnership Project (3GPP), Sophia-Antipolis, France. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/26_series/26.912/

[24] (2003, Sept.) Packet switched conversational multimedia applications; transport protocols (TS 26.236). Technical Specification Group Services and System Aspects, 3rd Generation Partnership Project (3GPP), Sophia-Antipolis, France. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/26_series/26.236/

[25] (2003, Oct.) Transparent end-to-end packet switched streaming service (PSS); RTP usage model (TR 26.937). Technical Specification Group Services and System Aspects, 3rd Generation Partnership Project (3GPP), Sophia-Antipolis, France. [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/26_series/26.937/

[26] (2002, June) Codec for circuit switched multimedia telephony service; general description (TS 26.110). Technical Specification Group Services and System Aspects, 3rd Generation Partnership Project (3GPP), Sophia-Antipolis, France. [Online]. Available: http://www.3gpp.org/ftp/Specs/latest/Rel-5/26_series/26 110-500.zip

[27] (2002, Mar.) Transparent end-to-end packet-switched streaming service stage 1 (TS 22.233). Technical Specification Group Services and System Aspects, 3rd Generation Partnership Project (3GPP), Sophia-Antipolis, France. [Online]. Available: http://www.3gpp.org/ftp/Specs/latest/Rel-5/22_series/22 233-500.zip

[28] (2002, Dec.) Multimedia messaging service (MMS); stage 1 (TS 22.140). Technical Specification Group Services and System Aspects, 3rd Generation Partnership Project (3GPP), Sophia-Antipolis, France. [Online]. Available: http://www.3gpp.org/ftp/Specs/latest/Rel-5/22_series/22 140-540.zip

[29] N. Faerber, B. Girod, and J. Villasenor, "Extensions of ITU-T recommendation of H.324 for error-resilient video transmission," *IEEE Commun. Mag.*, vol. 36, no. 6, pp. 120–128, June 1998.

[30] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A transport protocol for real-time applications," Internet Eng. Task Force, Request for Comments 3550, July 2003.

[31] H. Schulzrinne, A. Rao, and R. Lanphier, "Real time streaming protocol (RTSP)," Internet Eng. Task Force, Request for Comments 2326, Apr. 1998.

[32] M. Handley and V. Jacobson, "SDP: Session description protocol," Internet Eng. Task Force, Request for Comments 2327, Apr. 1998.

[33] L.-A. Larzon, M. Degermark, S. Pink, L.-E. Jonsson, and G. Fairhurst, "The Lightweight User Datagram Protocol (UDP-lite)," Internet Eng. Task Force, Request for Comments 3828, Jul. 2004.

[34] H. Zheng and J. Boyce, "An improved UDP protocol for video transmission over internet-to-wireless networks," *IEEE Trans. Multimedia*, vol. 3, no. 3, pp. 356–365, Sept. 2001.

[35] T. Stockhammer, H. Jenkac, and C. Weiss, "Feedback and error protection strategies for wireless progressive video transmission," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 465–482, June 2002.

[36] J. Ott, S. Wenger, N. Sato, C. Burmeister, and J. Rey. (2003, June) IETF Internet draft: Extended RTP profile for RTCP-based feedback (RTP/AVPF) [Online]. Available: http://www.ietf.org/internet-drafts/draft-ietf-avt-rtcp-feedback-07.txt

[37] T. Friedman, R. Caceres, and A. Clark, "RTP control protocol extended reports (RTCP XR)," Internet Eng. Task Force, Request for Comments 3611, Nov. 2003.

[38] J. Rey, D. Leon, A. Miyazaki, V. Varsa, and R. Hakenberg. (2003, Aug.) RTP retransmission payload format (in progress) [Online]. Available: http://www.ietf.org/internet-drafts/draft-ietf-avt-rtp-retransmission-10.txt

[39] H. Inamura, G. Montenegro, R. Ludwig, A. Gurtov, and F. Khafizov, "TCP over second (2.5G) and third (3G) generation wireless networks," Internet Eng. Task Force, Request for Comments 3481, Feb. 2003.

[40] R. Talluri, "Error-resilient video coding in the ISO MPEG-4 standard," *IEEE Commun. Mag.*, vol. 1, no. 1, pp. 112–119, June 1998.

[41] B. Yan and K. Ng, "A survey on the techniques for the transport of MPEG-4 video over wireless networks," *IEEE Trans. Consumer Electronics*, vol. 48, no. 4, pp. 863–873, Nov. 2002.

[42] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, July 2003.

[43] S. Wenger, M. Hannuksela, T. Stockhammer, M. Westerlund, and D. Signer, "RTP payload format for H.264 video," Internet Eng. Task Force, IETF draft-ietf-avt-rtp-h264-03.txt, Oct. 2003.

[44] C. Bormann, C. Burmeister, M. Degermark, H. Fukushima, H. Hannu, L.-E. Jonsson, R. Hakeberg, T. Koren, K. Le, Z. Liu, A. Martensson, A. Miyazaki, K. Svanbro, T. Wiebke, T. Yoshimura, and Z. H. Zhang, "Robust header compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed," Internet Eng. Task Force, Request for Comments 3095, July 2001.

[45] J. Zhang, J. Arnold, M. Frater, and M. Pickering, "Video error concealment using decoder motion vector estimation," in *Proc. IEEE TENCON'97*, vol. 2, pp. 777–780.

[46] D. Turaga and T. Chen, "Model-based error concealment for wireless video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 483–495, June 2002.

[47] B. Girod and N. Faerber, "Feedback-based error control for mobile video transmission," *Proc. IEEE*, vol. 87, no. 10, pp. 1707–1723, Oct. 1999.

[48] M. Wada, "Selective recovery of video packet loss using error concealment," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 807–814, June 1989.

[49] Y. Shan and A. Zakhor, "Cross layer techniques for adaptive video streaming over wireless networks," in *Proc. Int. Conf. Multimedia and Expo'02*, vol. 1, pp. 277–280.

[50] J. Kim, R. Mersereau, and Y. Altunbasak, "Error-resilient image and video transmission over the Internet using unequal error protection," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 121–131, Feb..

[51] M. van der Schaar and H. Radha, "Adaptive motion-compensation fine-granular-scalability (AMC-FGS) for wireless video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 360–371, June 2002.

[52] A. E. Mohr, E. A. Riskin, and R. E. Ladner, "Unequal loss protection: Graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 819–828, June 2000.

[53] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 511–523, June 2002.

[54] B. Girod, M. Kalman, N. Liang, and R. Zhang, "Advances in channel-adaptive video streaming," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 2002, pp. 9–12.

[55] M. Kalman, E. Steinbach, and B. Girod, "R-D optimized media streaming enhanced with adaptive media playout," in *Proc. IEEE Int. Conf. Multimedia and Expo 2002*, vol. 1, pp. 869–872.

[56] P. A. Chou and Z. Miao, "Rate-distortion optimized sender-driven streaming over best-effort networks," in *Proc. 4th Workshop IEEE Multimedia Signal Processing*, 2001, pp. 587–592.

[57] S. Wee, W. Tan, J. Apostolopoulos, and M. Etoh, "Optimized video streaming for networks with varying delay," in *Proc. IEEE Int. Conf. Multimedia and Expo 2002*, vol. 2, pp. 89–92.

[58] Q. Zhang, Z. Ji, W. Zhu, and Y.-Q. Zhang, "Power-minimized bit allocation for video communication over wireless channels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 398–410, June 2002.

[59] I.-M. Kim, H.-M. Kim, and D. Sachs, "Power-distortion optimized mode selection for transmission of VBR videos in CDMA systems," *IEEE Trans. Commun.*, vol. 51, no. 4, pp. 525–529, Apr. 2003.

[60] X. Lu, Y. Wang, and E. Erkip, "Power efficient H.263 video transmission over wireless channels," in *Proc. IEEE Int. Conf. Image Processing 2002*, vol. 1, pp. 533–536.

[61] Y. Eisenberg, C. Luna, T. Pappas, R. Berry, and A. Katsaggelos, "Joint source coding and transmission power management for energy efficient wireless video communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 411–424, June 2002.

[62] J. Cabrera, A. Ortega, and J. Ronda, "Stochastic rate-control of video coders for wireless channels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 496–510, June 2002.

[63] S. Aramvith, I.-M. Pao, and M.-T. Sun, "A rate-control scheme for video transport over wireless channels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 5, pp. 569–580, May 2001.

[64] G. De Los Reyes, A. Reibman, S.-F. Chang, and J.-I. Chuang, "Error-resilient transcoding for video over wireless channels," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 1063–1074, June 2000.

[65] H. Balakrishnan, V. N. Padmanabhan, S. Seshan, and R. Katz, "A comparison of mechanisms for improving TCP performance over wireless links," in *Proc. SIGCOMM'96*, pp. 256–269.

[66] A. Bakre and B. Badrinath, "Handoff and systems support for indirect TCP/IP," in *Proc. 2nd Symp. Mobile and Location-Independent Computing (MLICS'95)*, pp. 11–24.

[67] H. Balakrishnan and R. H. Katz, "Explicit loss notification and wireless Web performance," presented at the GLOBECOM'98, Sydney, Australia.

[68] T. Yoshimura, T. Ohya, T. Kawahara, and M. Etoh, "Rate and robustness control with RTP monitoring agent for mobile multimedia streaming," in *Proc. IEEE Int. Conf. Communications 2002*, vol. 4, 2002, pp. 2513–2517.

[69] Q. Wang and M. Abu-Rgheff, "Cross-layer signaling for next-generation wireless systems," in *Proc. IEEE Wireless Communications and Networking Conf. 2003*, vol. 2, 2003, pp. 1084–1089.

[70] M. Handley, S. Floyd, J. Padhye, and J. Widmer, "TCP friendly rate control (TFRC): Protocol specification," , Jan. 2003, Request for Comments 3448.

Minoru Etoh (Member, IEEE) received the B.E. and M.S.E.E. degrees from Hiroshima University, Hiroshima, Japan, in 1983 and 1985, respectively, and the Ph.D. degree from Osaka University, Osaka, Japan, in 1993.

He started his career as a Research Engineer at the Central Research Laboratories of Matsushita Electric in 1985. In the 1990s, he led an image communication research team and participated in MPEG-4 standardization. In May 2000, he joined NTT DoCoMo, Yokosuka, Japan. He was Director of Signal Processing at NTT DoCoMo Laboratory, at which he was involved in multimedia communication research and development. In 2002, he was appointed as President and CEO of DoCoMo Communications Laboratories USA, San Jose, CA, where he is now conducting several research groups in charge of mobile network architecture, security, terminal software, and mobile media technologies. He is also a Visiting Professor with Nara Institute of Science and Technology, Ikoma, Japan.

Takeshi Yoshimura (Member, IEEE) received the B.E. and M.E. degrees from the Department of Information and Communication Engineering, University of Tokyo, Tokyo, Japan, in 1997 and 1999, respectively.

He joined NTT Mobile Communication Network, Inc., Yokosuka, Japan, in 1999. Since then, he has been engaged in research on mobile streaming media technology, and participated in Internet Engineering Task Force and 3GPP standardization. He is currently a Research Engineer with NTT DoCoMo Multimedia Laboratories. His current research interests include audio signal processing and ultrasound communication.