3G Wireless Multimedia: Technologies and Practical Issues

Wenjun Zeng and Jiangtao Wen

PacketVideo Corporation
10350 Science Center Drive
San Diego, CA 92121
wzeng@packetvideo.com

ABSTRACT

This paper provides an overview of the emerging wireless communication standards, end-to-end wireless streaming systems, and relevant wireless multimedia technologies. It highlights some of the challenges in the deployment of 3G wireless multimedia services, using PacketVideo's solutions as an example.

1 INTRODUCTION

For years, the international community has anticipated the arrival of products to enable visual communication and rich multimedia experiences anywhere and at anytime. This anticipation is becoming a reality thanks to the International Telecommunication Union's (ITU) International Mobile Telecommunications "IMT-2000" initiative, also known as the Third Generation (3G) Mobile Systems. IMT-2000 is a single family of compatible standards that is intended to support both packet-switched and circuit-switched wireless data transmission and offer high data rates up to 2 Mbps (depending on mobility/velocity), with high spectrum efficiency. The 3G is typified by the convergence of voice and data with mobile Internet access, multimedia applications and high data transmission rates.

NTT DoCoMo, Japan's largest wireless operator, launched the world's first fully commercial 3G services on its W-CDMA (Wideband Code Division Multiple Access) based 3G system FOMA (Freedom of Mobile

Multimedia Access) on October 1, 2001. With high transmission speeds of up to 384 kilobits per second, the system can carry a video service referred to as "i-motion" that is an extension of the popular "i-mode" service [1]. Several US and European wireless operators have also unveiled their plans to deploy 2.5G (General Packet Radio Service (GPRS) or CDMA2000 1X) or 3G (Enhanced Data Rates for GSM Evolution (EDGE), W-CDMA, CDMA2000 1xEV-DO, etc.) networks in 2002 or 2003. Many of the major handset manufacturers have offered or are expected to offer 3G handsets in 2002. With the development of the 3G network infrastructure, emerging wireless communication standards and products, wireless multimedia applications and services are imminent and are poised to significantly change the way we live.

This paper provides an overview of the emerging wireless communication standards, end-to-end wireless streaming systems, and relevant wireless multimedia technologies. We will highlight some of the challenges in the deployment of 3G wireless multimedia services, using PacketVideo's solutions [2] as an example.

2 WIRELESS STREAMING STANDARDS

Interoperability is a crucial requirement in the telecommunication industry. It is a common belief that technology should be designed to permit maximum interoperability between implementations such that the marketplace for wireless video and next generation video services can evolve freely and openly. One of the key enablers to interoperability is the existence of open standard recommendations. In the case of wireless streaming multimedia, there are a number of well-developed component standards that cover video and audio compression, and packet-based communication. The emergence of a complete open standard for a streaming system is a relatively recent development. Some of the most important standards for wireless streaming are briefly described here.

2.1 Video and Audio Coding Standards

2.1.1 MPEG-4 Visual

The MPEG-4 video compression standard, known as MPEG-4 Visual [3], is quite broad in terms of bitrates and functionalities supported. It includes the support for H.263 baseline [4]. Some of the most important functionalities are scalability, error resilience, and object-based coding. Scalability brings the concept of encoding one time, and delivering the encoded bitstream to multiple decoders having different capabilities. Additionally, scalability permits adaptively changing the content bandwidth delivered from a server to a client. MPEG-4 facilitates these capabilities by including temporal enhancement frames, which increase the frame rate, and spatial enhancement frames, which increase the spatial quality and resolution of the frames.

MPEG-4 Visual can be separated into Profiles and Levels. The most prevalent use of MPEG-4 for wireless applications centers around the *Simple Visual Profile* and the *Simple Scalable Visual Profile* because of the computational resources and data rates associated with wireless handsets today. The *Simple Visual Profile* provides efficient, error resilient coding of rectangular video objects. The *Simple Scalable Visual Profile* adds support for coding of temporal and spatial scalable objects to the *Simple Visual Profile*. These are especially appropriate for delivering video over 2.5G and 3G wireless networks, which have highly variable throughput and are error-prone.

Recently a new video coding standard called MPEG-4 AVC (Advanced Video Codec) or ITU H.264 emerges [5]. It is a joint effort between two standards bodies, i.e., ISO/IEC JTC1 and ITU-T. The main goal of this standard has been to improve the coding efficiency, at possibly a cost in increase of computational complexity and memory requirements. For embedded systems, the cost could be significant. Current estimates are that MPEG-4 AVC demands about 2-2.5x factor in complexity over H.263+ Streaming Wireless Profile and MPEG-4 Simple Visual Profile, and 2.5-3x factor in complexity over H.263 baseline. Furthermore, if a full implementation of the encoder is implemented (i.e., with multiple reference frames for motion estimation), the complexity can increase to as much as 10 times that of H.263 baseline. In addition, because of the increase in the number of reference frames that must be supported in MPEG-4 AVC (minimum of 3), additional frame buffers must be allocated in the decoder. Therefore the impact of MPEG-4 AVC on embedded devices needs to be carefully evaluated.

2.1.2 MPEG-4 Systems

MPEG-4 Systems [6] consists of several key components for transmitting and communicating MPEG-4 audio and visual information for different purposes. It covers scene description for the presentation of audio-visual objects, an object description framework, intellectual property rights management and protection (IPMP) framework, demultiplexing of multiple compressed bitstreams, management of the receiving terminal's buffers, and time identification.

One of the components of MPEG-4 Systems is the MP4 File Format, which is designed for the storage and streaming of MPEG-4 audio and visual information. It includes an optional hint track, with information that helps a server parse the data portion of the file for efficient delivery. MP4 File Format has been recently adopted by 3GPP, the 3rd Generation Partnership Project [7], as the file format for the MMS (Multimedia Messaging Service) standard.

Another important component of MPEG-4 system is IPMP extensions [8]. MPEG is developing a framework and specifications for extending the IPMP capabilities in MPEG. It is aimed at allowing the same IPMP protected content to be consumed on MPEG IPMP compliant devices from different vendors, and the same content to be protected by different vendors' standard compliant tools.

2.1.3 Speech and Audio Coding Standards

The focus of the wireless multimedia standardization community has been on existing speech codecs for cellular communications. A number of codecs outside of MPEG-4 have been adopted by implementers for real time communication over wireless networks. Two such codecs are AMR (Adaptive Multi-Rate) [9] supported by the GSM (Global System for Mobile Communication) and 3GPP community [7] and EVRC (Enhanced Variable Rate Coder) [10] supported by CDMA and 3GPP2 community [11], both have good error resiliency. MPEG-4 audio includes several audio codecs, including MPEG-4 AAC (Advanced Audio Coding), to support higher bandwidth and scalable audio communication.

2.2 Wireless Multimedia Forum (WMF)

The WMF was started in early 2000. At that time, neither 3GPP nor 3GPP2 were focused on video streaming. The goals of this forum included establishing technology consensus around a set of protocols that are suitable for use in streaming multimedia over a wireless network. The consensus of technical experts from a group of companies that are recognized as market leaders in the content, wireless distribution, and terminal equipment industries has been used to identify this set of streaming protocols. In Jaunary 2001, the WMF succeeded in producing a key deliverable: the first set of identified protocols documented in the Recommended Technical Framework Document (RTFD) Version 1.0 [12] which serves to facilitate and guide other standard bodes' recommendations for wireless streaming. After two important years helping foster the mobile multimedia industry, which resulted in the adoption of many of the WMF recommendations by 3GPP [7] and 3GPP2 [11], the WMF has fulfilled its mission and no longer exists as an organization.

2.3 3GPP and 3GPP2

One example of convergence in standards driven by WMF is the fact that both 3GPP [7] and 3GPP2 [11], the 3rd Generation Partnership Projects which are the most important organizations for defining specifications for use by 3G wireless carriers and terminal manufactures, have recommended specifications for streaming service which are subsets of the WMF RTFD 1.0 with respect to video and speech codecs. The scope of 3GPP includes the maintenance and development of the GSM technical specifications and technical reports including evolved radio access technologies (e.g. GPRS and EDGE). 3GPP2 is a parallel Partnership Project responsible for the development of the detailed specifications of CDMA Multi Carrier member of the IMT-2000 family. WMF and the 3GPPs have worked together and influenced one another to guarantee that the specifications developed by these bodes are harmonized.

Among others, 3G-324M is a standard way developed by 3GPP to enable video communication over 3G, and to ensure interoperability between landline video systems and video-enabled mobile devices [13], It is closely tied to the ITU-T H.324 standard for wireline videoconferencing [14], and was designed to operate via the wireless and mobile network of UMTS (Universal Mobile Telecommunications System) over 64kbit/s circuit-switched lines. It includes H.223 [15] as the multiplex protocol, H.245 [16] as the control protocol, H.263 and MPEG-4 simple profile level 1 as the default video codecs, GSM-AMR and G.723.1 [17] as the default speech codecs. In fact, a similar group of protocols have been adopted by 3GPP2 for circuit-switched video conferencing services in 3GPP2 wireless telecommunications networks [18].

Recently, 3GPP adopted a framework for Internet Protocol (IP) based streaming applications in 3G networks [19], where protocols for control signaling, scene description, media transport and media encapsulations are specified, and codecs for speech, audio, video, still images, bitmap graphics, and text are specified. A significant subset of this framework is reflected in Section 3 where an overview of a typical wireless streaming multimedia system is presented.

2.4 IETF Streaming Protocols

The Internet Engineering Task Force (IETF) is a voluntary standards body that is dedicated to making recommendations for how to communicate information over the IP networks. The IETF has recommended a number of methods based on RTP (Real Time Protocol) [20] and RTSP (Real Time Streaming Protocol) [21] for the delivery of video bitstreams with synchronized audio from a server to a terminal. The protocol stack for the delivery of multimedia data in a packet-based environment is shown in Fig. 1.

RTP is an application layer component that utilizes UDP (User Datagram Protocol), a "best-effort", connectionless protocol, as a transport mechanism. It is therefore suitable for delivering data that must arrive without delay. RTP includes a sub-component known as RTCP (Real Time Control Protocol) that is used to control performance information between a server and a client. RTSP is a session-oriented protocol that is reliably transported over TCP (Transmission Control Protocol) between the server and the client. It uses SDP (Session Description Protocol) [22] to carry all descriptive information associated with the streaming session. Some of the popular recommendations from IETF for video streaming include RFC3016, the RTP payload format for MPEG4 video and audio [23], and RFC2429, the RTP payload format for H.263 [24], both are adopted by 3GPP for IP based packet switched networks.

3 WIRELESS STREAMING SYSTEM OVERVIEW

Fig. 2 shows an overview diagram of a typical wireless streaming multimedia system. It can be viewed as a set of protocols, working in concert to deliver wireless video and audio. It comprises three major subsystems, i.e., content creation subsystem, multimedia distribution subsystem, and multimedia terminal subsystem. Each major subsystem is divided into mandatory features that are needed to guarantee interoperability in the network, and optional features that will lead to better performance and user experience in a deployed system.

All the mandatory features and most of the optional features are defined in a formal international standard. As a whole, the system illustrated herein is fully compliant with WMF RTFD 1.0 and can be used to generate a system to interoperate with the 3GPP and 3GPP2 recommended streaming multimedia terminals.

In the next section, we use PacketVideo's technologies as an example to highlight some of the most challenging issues in the deployment of a wireless multimedia streaming system.

4 PACKETVIDEO'S TECHNOLOGIES

PacketVideo has developed technologies specifically for wireless multimedia applications [25]. The technologies can be roughly categorized into two categories: algorithmic optimization and implementation/architectural optimization.

4.1 Algorithmic Optimization

For an end-to-end wireless multimedia streaming system, algorithmic optimization includes encoder optimization, decoder quality optimization, and joint source-network optimization.

4.1.1 Encoder Optimization

For content creation, a number of signal processing functions such as source noise pre-filtering, scene change detection, and rate control are critical for achieving good quality of the compressed video. Among others, rate control is a central piece for video codecs to achieve consistent good quality across the whole sequence under the channel bandwidth and buffer constraints. One of the challenges in rate control for wireless multimedia applications is to be able to handle video sequences of varying characteristics in a simple and coherent way. Most existing rate control solutions suffer from two fundamental problems, i.e., stationarity assumption which results in constant bit allocation among GOPs (Group of Pictures), and model parameter mis-estimation. These prior solutions, although may work fine for some standard test sequences that are usually simple and more stationary, typically have problems dealing with more complex sequences. We have developed a sequence based bit allocation framework with the capability of tracking the non-stationary characteristics in the video source without resorting to look-ahead encoding [26][27]. In particular, a R-mad (rate vs. mean absolute difference) model is proposed to address bit allocation among all frames in a video sequence (which implicitly addresses scene change detection, and bit allocations for *I*, *P*, *B* frames) in a coherent way, and a general concept of bit allocation guarantee is presented to achieve the allocated bits in a deterministic way. Such rate control solution makes full use of the decoder buffer to achieve constant quality video with less quality flicker and motion jerkiness, and is applicable for real time applications. It also provides the functionality of proactively adjusting the actual frame rate in return for higher image quality in a more deterministic way, as opposed to the passive random bursty frame dropping experienced in many prior methods.

Fig. 3 shows a comparison of the proposed rate control solution with the MPEG-4 Annex L rate control solution [3]. It can be seen that our proposed solution can achieve much less PSNR variation and frame dropping than MPEG-4 Annex L rate control, which leads to more consistent visual quality across the sequence. On average, the PSNR gain is 0.86 dB. Note that when a frame was skipped, the previous encoded frame was used in the PSNR calculation based on the fact that the decoder displays the previous encoded frame instead. As a result, the skipped frames can usually be easily identified, e.g., those points with lower than 25 dB PSNR value. In particular, the proposed solution achieves much better quality than MPEG-4 Annex L in the scene change frames of the sequence, thus significantly reducing the annoying flickering effect.

In addition to rate control, fast rate-distortion optimized coding mode selection and motion estimation have also been developed to further improve the video quality. All these different quality optimization modules in PacketVideo’s encoder result in one of the best MPEG-4 compliant encoder solutions in the market.

The MPEG-4 visual specification provides error detection, localization and robust entropy coding tools to improve the bitstream's resilience to transmissions errors. Bit errors in the received bitstream are localized through the use of resynchronization markers. The resynchronization markers bound the data into "video packets", with packet header inserted, which allows independent decoding of each packet and prevents error propagation between packets. One tool within the MPEG-4 standard that is especially useful for localization is "data partitioning". With this syntax, the motion information is separated from the texture information, which facilitates the recovery of the more important motion vector information in the case of bit errors, as will be discussed in the next sub-section. "Intra-macroblock (MB) refreshing" by forcing a few MBs to be intra-coded in each inter-predictive coded frame also helps prevent temporal error propagation. As a result, existing MPEG4 contents may have to be re-purposed to meet the requirements of wireless delivery.

4.1.2 Decoder Quality Optimization

The decoder components are equally important to the generation of good quality video on handheld 3G terminals, especially for video delivery over error prone wireless channels.

In MPEG-4, the implementer has control over several key components in a video decoder that are outside the scope of the standard. PacketVideo differentiates its implementation of the MPEG-4 decoder by implementing special patent pending algorithms in the Bitstream Decoder and in the Post-Processor. In the Bitstream Decoder, the detection, localization, containment and concealment of errors is one of the most crucial components of PacketVideo technology for wireless video. PacketVideo algorithms also address compression artifacts with practical solutions for the complex post-processing problem, which is critical to the subjective quality of the decoded video, especially at lower bitrates.

As mentioned before, one tool within the MPEG-4 standard that is especially useful for localization is "data partitioning". With this syntax, when motion vectors are correctly reconstructed but errors are detected in the texture component, it is possible to use only the motion vector information to re-create the decoded image.

When errors are detected in the motion data, PacketVideo algorithms recover lost motion vector information via smart bitstream decoding [28], which is then used for concealment. This “motion-compensated concealment” approach results in decoded video that hides serious artifacts from the viewer. This PacketVideo error resilience technology is known as SignalTrack™ [2].

Fig. 4 shows the effects of burst errors (W-CDMA errors at 10^{-3} BER on average) applied to an MPEG-4 simple profile video. PacketVideo SignalTrack™ algorithms are used on the right to recover and conceal the errors, while the decoder on the left simply resynchronizes after errors, and continues to decode. This clearly shows the benefit of PacketVideo error resilient decoding techniques.

4.1.3 Joint Source-network Processing

In addition to error resilience, wireless multimedia delivery faces several other challenges, especially in dealing with bandwidth variations [29]. Wireless channel bandwidth can vary significantly, depending on the signal strength and interference level that a user receives. As a result, as a user travels through different parts of the cell, different bandwidths may be dynamically assigned to the user. In addition, depending on the quality of service (QoS) capability of the wireless network, multi-user sharing of the wireless channel with heterogeneous data types can also lead to significant user channel bandwidth variation. This bandwidth variation can further lead to network buffer overflow and hence packet loss. Finally, data transmission can be interrupted completely due to, e.g., cell reselection/handoff process, resulting in transmission gaps ranging from a fraction of a second to several seconds. This unpredictability of available wireless channel bandwidth introduces high delay jitter for the multimedia streaming data.

QoS guarantee in a wireless network is often costly, if not impossible, since the capacity of a wireless network is related to the uncontrollable RF condition. To deal with such issues, technologies such as dynamic rate control (DRC) based on estimated network bandwidth [25][29] selective retransmission [30], and streaming in advance [31], need to be employed.

PacketVideo's FrameTrack™ technology [2] deals with limited or soft QoS provisioning environments and provides the ability to dynamically manage wireless media traffic on a network or per subscriber basis, thereby provides good video quality even in the error-prone wireless environment. It does so by making use of MPEG-4 temporal scalability provision and by automatically and dynamically adjusting the streaming bit rate in the presence of channel fluctuation through DRC to ensure the best user experience for each customer.

FrameTrack™ is an end-to-end joint source-network-terminal processing process. Figure 4 illustrates FrameTrack™ with RTCP feedback control. First, content is encoded with PVAuthor, PacketVideo's MPEG4 encoder, using MPEG-4 temporal scalability. During streaming, PVPlayer monitors many different aspects of the PacketVideo stream to detect packet loss and fluctuations in bandwidth and communicate them through RTCP reports to the streaming module. The streaming module then uses a unique set of algorithms, specifically designed for wireless networks, to makes a rate decision and regulate the frame rate. The result is uninterrupted video streaming at decent quality. FrameTrack™ also provides enhanced bandwidth management, including scaling to support many ranges of data rates. With FrameTrack™, one encoded MPEG-4 file can be used to deliver multiple streams at multiple bit rates to multiple devices. It has been demonstrated that PacketVideo DRC streaming server effectively tracks the variation of bandwidth for scalable video streaming and significantly reduces player rebuffering [29].

Another way to deal with channel bandwidth variation is, albeit at the cost of storage overhead, to do dynamic bitstream switching among multiple bitstreams with different bit rates [30]. PacketVideo has developed fast intelligent switching algorithms to reduce the drifting artifacts caused by bitstream switching. With such intelligent switching, more complex solution such as the SP frame solution [32] may not deem necessary. The general concept of rate-distortion optimized streaming [33] can also be exploited to further improve the performance.

To facilitate efficient network and service planning, it is important to have a good understanding of the feasibility and impact of delivering streaming video over a 3G network. Packetvideo has performed various field trials and lab trials on streaming video over various 2.5G and 3G networks. For example, both mixed traffic wireless system performance and video streaming performance have been investigated using an event-driven system-level simulation and a point-to-point CDMA2000 1x emulator [29]. It has been found that the average throughput for scalable video streaming is much better than that for non-scalable video streaming under a pure video traffic condition. Moreover, the packet loss rate for scalable video streaming is much lower than that for non-scalable video streaming, which demonstrates that scalability is a critical element for a successful deployment of wireless video streaming service. It was also found that with a mixed scalable video and web traffic, the CDMA2000 1x system is able to achieve higher spectrum efficiency than that for web traffic only. The effects of radio link protocol (RLP) retransmissions have also been investigated using a CDMA2000 1x emulator.

Other approaches to deal with channel bandwidth variation and packet loss include selective retransmission [30] and streaming in advance [31]. In selective retransmission, packets can be prioritized so that only critical packets, when lost, are resent in an effort to conserve bandwidth. Selective retransmission can help achieve better overall throughput and quality on bursty networks. Streaming in advance refers to the technique of streaming faster than real time (i.e., playback time) when there is additional bandwidth available on the network. This will potentially allow the initial pre-roll buffer to fill faster than real time, allowing playback to start quicker. It also allows a better handling of the channel bandwidth variation so that the playback can be smoother.

4.1.4 Format Compliant Selective Encryption and Shuffling

One important component in the MPEG IPMP system is access control. One of the major goals of content access control for entertainment purposes (as opposed to for top secret communications, e.g. of military information) is to enable authorized users to view the video, and to disallow unauthorized users to view the video with satisfactory quality, similar to a set top cable TV scrambler. One common method for access control is through encryption and/or shuffling. However, the traditional bitstream-structure-agnostic wholesale encryption approach poses significant problems handling challenges in today's content production to consumption food chain, especially in the capability of providing a flexible complexity/security tradeoff, network friendliness, and error resiliency of the protected bitstream, which are critical for wireless delivery. In addition to the above challenges, there are other desirable features that most multimedia applications have in common, such as applicability of various signal processing such as compressed domain watermarking, random access, scene change detection, content-based searching or filtering, etc. to the encrypted bitstream (without decryption) in the various links of an end-to-end chain.

To meet these requirements, PacketVideo has developed a format compliant selective encryption and shuffling framework for access control of compressed content [34][35][36]. The framework has been recently adopted into MPEG4 IPMP extensions [8]. It is parametrically configurable and is capable of maintaining full bitstream format compliance (therefore inheriting the important properties of scalability and error resilience of the standards) while providing the desired complexity/security trade off suitable for the content, platform, network, and application. Legacy encryption schemes such as wholesale encryption or bitstream structure agnostic run-length based encryption can also be supported.

4.2 Implementation Optimization

At the heart of PacketVideo solutions are optimal implementations of standards and algorithms for audio, video and communication. One of the key differentiators of PacketVideo technology is its ability to effectively run on many of the world's leading wireless platforms and operating systems, including ARM, Intel, Lucent, Texas Instrument, Qualcomm, Infineon, Symbian EPOC, WinCE and others. PacketVideo is also working in cooperation with the world's leading handset manufacturers such as Casio, Compaq, Motorola, Sanyo, Sendo, Sharp, Fujitsu, and many others in order to integrate PacketVideo technology into multimedia-capable mobile phones and PDAs. In addition to algorithmic optimization, PacketVideo has been developing software-based solutions that focus on core software architecture optimization, C source code optimization, RTOS (real time operating system)-dependent optimization, and processor-specific optimization. Each of these stages of optimization leads to a final product in the most appropriate form for the various hardware platforms used in the wireless multimedia space.

In particular, PacketVideo technology focuses on developing optimal software libraries that can be used to implement encoders and decoders on many embedded platforms with limited memory and processing power. The computational requirements (MIPS) and memory (both program and data) efficiency are very important. Modular design is especially important for software reuse. Potential computationally expensive parts are made modular so that they can be optimized for specific processors, platforms, and RTOS's. For example, computationally complex operations in a typical video codec include block matching for motion estimation, motion compensation, discrete cosine transform, interpolation, color conversion, etc. In addition, for embedded digital signal processors, memory access to external memory is often a significant bottleneck in the implementation of real-time embedded video systems that have large memory requirements. Given the diversity of the power of different communication terminals, it is sometimes desirable that a codec can provide complexity scalability.

For commercial deployment, it is important for the streaming system to support large-scale, interoperable, distributed, and robust streaming service. To that end, technologies such as splitters, high availability encoding solution, load balancing, network-shared storage etc. have been developed. Splitters are intermediate servers that re-broadcast the incoming packets to their respective clients [30]. High availability encoding solution refers to encoders that support connections to many servers simultaneously. Load balancing generally refers to the division of processes, normally handled by a single server, among two or more servers. The combination of all these different approaches will result in a more robust, large-scale, scalable, and distributed streaming system.

With respect to MPEG-4 delivery, Fig. 5 shows three of PacketVideo's products, and their interaction. It comprises pvAuthor™ for encoding content, pvServer™ for controlling and delivering rich media, and pvPlayer™ for decoding and displaying the content on a mobile device. Not shown are the PVServer infrastructure service features optimized for multimedia delivery in a wireless carrier's network. These features include capabilities such as subscriber provisioning and authorization, billing interfaces, and server event monitoring. The system is compliant with WMF/3GPP and is built around PacketVideo's QualityTrack™ technology, which include features such as DeviceTrack™, SoundTrack™, FastTrack Download™, FrameTrack™ and SignalTrack™ [2].

Despite various standardization efforts, there still exist discrepancies between different proprietary systems, between proprietary and standard-compliant systems, and between different standard bodies' recommendations. To enable interoperability between the increasingly diverse devices on the markets, transcoding servers and gateway technologies have been developed that translate content to suitable formats for various platforms, regardless of protocol, application, screen size, and language used. For example, PacketVideo and NTT DoCoMo have jointly developed a gateway technology that enables MPEG-4 streams transferred in RTP from PVServer™ to be converted into a 3G-324M-based format compliant with NTT DoCoMo's FOMA video-enabled phones [1]. This enables the delivery of a single live stream to numerous handsets simultaneously, whereas current 3G-324M videophone services only allow one-to-one communication. This combination expands potential applications from pure video telephony to a wide variety of compelling live and pre-encoded information, communication and entertainment applications.

5. CONCLUSION

This paper highlights that, through a strong concentration on the use of international standards, wireless multimedia technology may be used successfully to interoperate between different vendors' solutions. An overview of the end-to-end wireless streaming systems is given, and some of the challenges in the deployment of 3G wireless multimedia services, including encoder quality optimization, bandwidth scalability, error resiliency, post-processing, joint source-network optimization, implementation optimization etc., are discussed.

6. ACKNOWLEDGEMENT

The authors would like to thank Mark Banham, Joe Huang, Bo Xie and Szu-wei Wang for providing supporting materials and valuable comments.

7. REFERENCES

[1] NTT DoCoMo website: http://www.nttdocomo.com/top.html

[2] PacketVideo website: http://www.pv.com

[3] ISO/IEC 14496-2:1999, "Information technology - Coding of audio-visual objects - Part2: Visual", December 1999.

[4] "Draft ITU-T Recommendation H.263." May 2, 1996.

[5] Joint Video Team (JVT) of ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, "Joint Committee Draft," May 2002.

[6] ISO/IEC 14496-1:1999, "Information technology - Coding of audio-visual objects - Part1: Systems", December 1999.

[7] The 3rd Generation Partnership Project (3GPP), http://www.3gpp.org/

[8] MPEG4 IPMP FPDAM, ISO/IEC 14496-1:2001/ AMD3, *ISO/IEC JTC 1/SC 29/WG11* N4701, March 2002.

[9] GSM 06.90 “Digital Cellular Telecommunication System; Adaptive Multi-Rate (AMR) Speech Transcoding”.

[10] 3GPP2, “Enhanced variable rate codec (EVRC),” document C.S0014-0

[11] The 3rd Generation Partnership Project 2 (3GPP2), http://www.3gpp2.org/

[12] Wireless Multimedia Forum, Technical Working Group, “Recommended Technical Framework Document, Version 1.0”, January 2001

[13] 3G TR 26.911 v 1.1.0 (1999) 3rd Generation Partnership Project, TSG-SA Codec Working Group, “Codec(s) for Circuit Switched Multimedia Telephony Service, Terminal Implementer’s Guide”.

[14] ITU-T Recommendation H.324 (1998) “Terminal For Low Bitrate Multimedia Communication”.

[15] ITU-T Recommendation H.223 (1996) “Multiplexing protocol for low bitrate multimedia communication”.

[16] ITU-T Recommendation H.245 (1999) “Control protocol for multimedia communication”.

[17] ITU-T Recommendation G.723. “Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 bit/s”.

[18] 3GPP2 C.P9016, “Circuit-switched video conferencing services,” version 0.99, April 2002.

[19] 3GPP TS 26.234, v 1.5.1 (2001-03), 3rd Generation Partnership Project, TSG-SA, “Transparent end-to-end packet switched streaming service, Protocols and codecs, Release 4”.

[20] IETF RFC: 1889- RTP: A Transport Protocol for Real Time Applications, H. Schulzrinne, et al, Jan. 1996.

[21] IETF RFC: 2326 – Real Time Streaming Protocol (RTSP), H. Schulzrinne, A. Rao, R. Lanphier, April 1998.

[22] IETF RFC: 2327 – Session Description Protocol (SDP), M. Handley, V. Jacobson, April 1998.

[23] Y. Kikuchi et al, “RTP Payload Format for MPEG-4 Audio/Visual Streams,” IETF RFC 3016, Nov. 2000.

[24] C. Bormann et al, “RTP Payload Format for the 1998 Version of ITU-T Rec. H.263 Video (H.263+),” IETF RFC2429, Oct. 1998.

[25] “PacketVideo Technology White Paper”, http://www.pv.com/prodtech/tech.asp

[26] B. Xie and W. Zeng, “Sequence-based rate control for constant quality video", in *Proc. Inter. Conf. Image Proc.*, Sept 2002.

[27] B. Xie and W. Zeng, “A Sequence-based rate control framework for constant quality video", submitted to *IEEE Trans. Cir. & Sys. for Video Tech.*, 2002.

[28] J. Wen, J. Vollasenor, "Soft-input soft-output decoding of variable length codes", *IEEE Trans. Com.* vol. 50, no. 5, pp 689 –692.

[29] J. Huang, Y. Yao, Y. Bai and S. Wang, "Scalable video streaming over a mixed-traffic CDMA2000 1x wireless network," to appear in *IEEE Trans. Cir. & Sys. for Video Tech.*, 2002.

[30] G. Conklin, G. Greenbaum, K. Lillevold, A. Lippman and Y. Reznik, "Video coding for streaming media delivery on the Internet," *IEEE Trans. Cir. & Sys. for Video Tech.*, vol. 11, no. 3, March 2001.

[31] Microsoft Windows Media "Corona", http://www.microsoft.com/windows/windowsmedia/thirdgen/default.asp.

[32] Marta Karczewicz and Ragip Kurceren, "A Proposal for SP-frames", document VCEG-L-27, ITU-T Video Coding Experts Group Meeting, Eibsee, Germany, 09-12 January 2001.

[33] P. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," submitted to *IEEE Trans. Multimedia*, Feb., 2001

[34] J. Wen, M. Severa, W. Zeng, M. Luttrell and W. Jin, "A format compliant configurable encryption framework for access control of multimedia," in *IEEE Workshop on Multimedia Signal Processing*, pp. 435-440, Oct. 2001.

[35] W. Zeng, J. Wen, and M. Severa, "Fast self-synchronous content scrambling by spatially shuffling codewords of compressed bitstreams," in *Proc. Inter. Conf. Image Proc.*, Sept 2002.

[36] J. Wen, M. Severa, W. Zeng, M. Luttrell and W. Jin, "A format compliant configurable encryption framework for access control of video," *IEEE Trans. Cir. & Sys. for Video Tech., Special Issue on Wireless Video*, pp. 545-557, June 2002.

<table>
<tr><td>Application Control Commands</td><td>Audio Data, Video Data,
Sender/Receiver Reports</td></tr>
<tr><td>RTSP</td><td>RTP/RTCP</td></tr>
<tr><td>TCP</td><td>UDP</td></tr>
<tr><td colspan="2">IP</td></tr>
<tr><td colspan="2">Radio Link/Data Link</td></tr>
<tr><td colspan="2">Physical Layer</td></tr>
</table>

Figure 1: Network protocol stack for delivering one-way multimedia data over the IP networks.

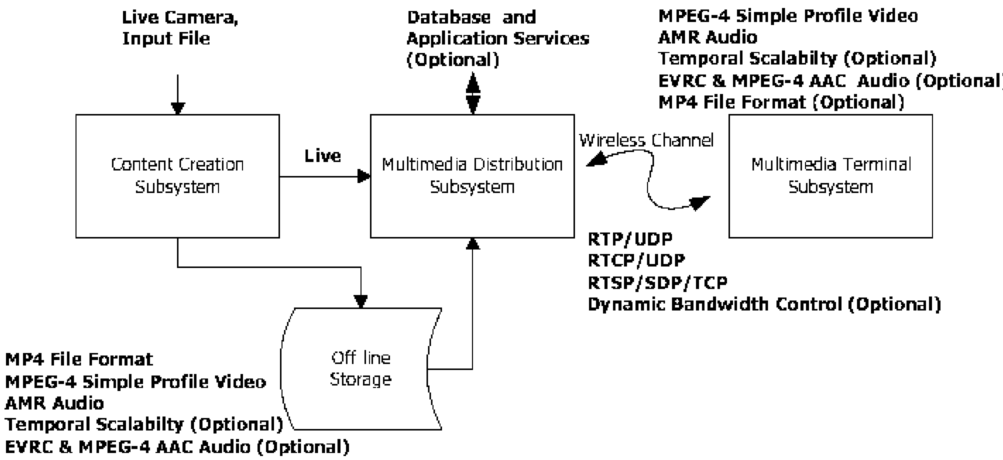

Fig. 2: Wireless streaming multimedia system overview.

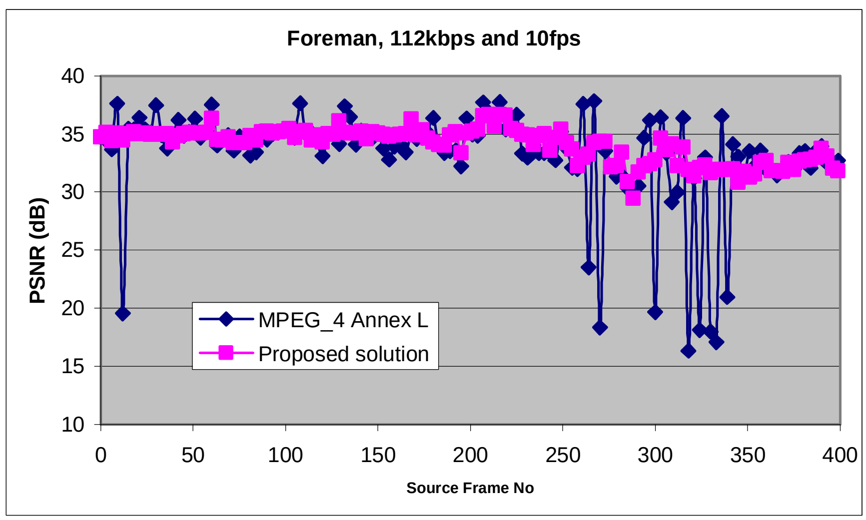

Fig. 3: PSNR variations for the “Foreman” sequence encoded at 112 kbps and 10 frames per second.

Fig. 4: Example of MPEG-4 decoding without PacketVideo error recovery and concealment (left) vs. PacketVideo error resilient decoding (right) in a 10^{-3} avg. BER W-CDMA channel.

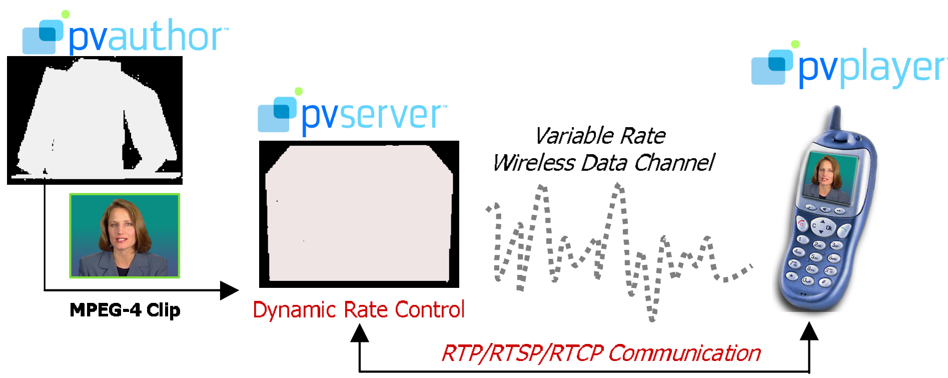

Fig. 5. PacketVideo products and their interaction.