1

Multimedia Networking and Communication: Principles and Challenges

Mihaela van der Schaar and Philip A. Chou

In case you haven't noticed, multimedia communication over IP and wireless networks is exploding. Applications such as BitTorrent, used primarily for video downloads, now take up the lion's share of all traffic on the Internet. Music file sharing, once on the legal cutting edge of massive copyright infringement on college campuses around the world, has moved into the mainstream with significant legal downloads of music and video to devices such as the iPod and numerous other portable media players. Multimedia podcasting to client computers and portable devices is a phenomenon exploding in its own right. Internet radio, pioneered in the late 1990s, is now being joined in a big way by peer-to-peer television such as CoolStreaming and PPLive. Audio and video on demand over the Internet, also available since the late 1990s on the Web sites of well-funded organizations such as CNN.com and MSNBC.com, are now at the core of a multitude of new music and video businesses from Napster to iTunes to MTV's Urge service, and will be expanding imminently to full-length movie delivery on demand. Moreover, Web sites such as YouTube have made publishing videos on demand available to anyone with a home video camera, which these days is nearly everyone with a mobile phone. Indeed, most mobile phones today can actively download *and upload* both photos and videos, sometimes in real time. Internet telephony is exploding, with popular applications such as Skype and others enabling wideband voice and video conferencing over the Internet. In general, voice over IP (VoIP) is revolutionizing the telecommunications industry, as circuit-switched equipment from PBX to long haul equipment is being replaced by soft IP switches. Enhanced television is also being delivered into the living room over IP networks by traditional telephone providers through DSL.

Once inside the home, consumer electronics manufacturers, and increasingly, the computer industry and its partners, are distributing audio and video over WiFi to monitors and speaker systems around the house. Now that the analog-to-digital revolution is nearly complete, we are undergoing an all-media-over-IP revolution, with radio, television, telephony, and stored media all currently being delivered over IP wireline and wireless networks. To top it all off, brand new types of media, such as game data for interactive gaming over the Internet, are strongly emerging.

Despite having unleashed a plethora of new multimedia applications, the Internet and wireless networks provide only limited support for multimedia. The Internet and wireless networks have inherently unpredictable and variable conditions. If averaged over time, this variability may not significantly impact delay-insensitive applications such as file transfer. However, variations in network conditions can have considerable consequences for real-time multimedia applications and can lead to unsatisfactory user experience. Multimedia applications tend to be *delay sensitive, bandwidth intense, and loss tolerant*. These properties can change the fundamental principles of communication design for these applications.

The concepts, theories, and solutions that have traditionally been taught in information theory, communication, and signal processing courses may not be directly applicable to highly time-varying channel conditions, adaptive and delay-sensitive multimedia applications, and interactive multiuser transmission environments. Consequently, in recent years, the area of multimedia communication and networking has emerged not only as a very active and challenging integrative research topic across the borders of signal processing and communication, but also as a core curriculum that requires its own set of fundamental concepts and algorithms that differ from those taught in conventional signal processing and communication courses.

This book aims at providing the reader with an in-depth understanding of the theoretical foundations, key design principles, algorithms, and existing standards for multimedia communication and networking.

This introductory chapter provides motivation for studying the topic of multimedia communication, the addressed applications, and associated challenges. Subsequently, a road map of the various chapters is provided. A suggested use for graduate instruction and self-study is also provided.

1.1 DIMENSIONS OF MULTIMEDIA COMMUNICATION

1.1.1 Multimedia Communication Applications

The emergence of communication infrastructures such as the Internet and wireless networks enabled the proliferation of the aforementioned multimedia applications. These applications range from simple music downloading to a portable device, to watching TV through the Internet on a laptop, or to viewing movie trailers posted on the Web via a wireless link. Some of these applications are new to the Internet revolution, while others may seem more traditional, such as sending VoIP to an apparently conventional telephone, sending television over IP to an apparently conventional set top box, or sending music over WiFi to an apparently conventional stereo amplifier.

An obvious question that comes to mind when considering all the aforementioned applications is how to jointly discuss these applications. What do they have in common and how do they differ? To provide an answer to this seemingly simple question, we will discuss the various dimensions of these multimedia communication applications.

1.1.2 Streaming Versus Downloading

Conventional downloading applications (e.g., file transfer such as FTP) involve downloading a file before it is viewed or consumed by a user. Examples of such multimedia downloading applications are downloading an MP3 song to a portable device, downloading a video file to a computer via BitTorrent, or downloading a podcast. (Despite its name, podcasting is a “pull” technology with which a Web site is periodically polled for new multimedia content.) Downloading is usually a very robust way to deliver media to a user. However, downloading has two potentially important disadvantages for multimedia applications. First, a large buffer is required whenever a large media file (e.g., an MPEG-4 movie) is downloaded. Second, the amount of time required for the download can be relatively large, thereby requiring the user to wait minutes or even hours before being able to consume the content. Thus, while downloading is simple and robust, it provides only limited flexibility both to users and to application designers.

An alternative to downloading is streaming. Streaming applications split the media bit stream into separate chunks (e.g., packets), which can be transmitted independently. This enables the receiver to decode and play back the parts of the bit stream that are already received. The transmitter continues to send multimedia data chunks while the receiver decodes and simultaneously plays back other, already received parts of the bit stream. This enables low delay between the moment data is sent by the transmitter to the moment it is viewed by the user. Low delay is of paramount importance for interactive applications such as video conferencing, but it is also important both for video on demand, where the user may desire to change channels or programs quickly, and for live broadcast, where the content length is unbounded a priori, but the delay must be finite. Another advantage of streaming is its relatively low storage requirements and increased flexibility for the user, compared to downloading. However, streaming applications, unlike downloading applications, have deadlines and other timing requirements to ensure continuous real-time media playout. This leads to new challenges for designing communication systems to best support multimedia streaming applications.

1.1.3 Streaming Media on Demand, Live Broadcast, and Real-Time Communication

Multimedia streaming applications can be partitioned into three classes by delay tolerance. Interactive audio and video telephony, teleconferencing, and gaming have extremely low delay tolerance, usually no more than 200 ms of end-to-end delay for comfortable interaction. In contrast, live broadcast applications (e.g., Internet radio), which typically have no interactivity, have a large delay tolerance, say up to 30 s, because the delay cannot be detected without interactivity and without a reference, such as a neighbor who is listening to a conventional radio. (Cheers coming from a neighbor's apartment 30 s before a goal can certainly ruin the surprise!) Intermediate in terms of delay tolerance is the application of streaming media on demand, which has only moderate interactivity requirements, such as channel changing and VCR-like control. The differences in delay tolerance among these three classes of multimedia applications have profound effects on their design, particularly with respect to error recovery. Low-delay, low bit rate applications such as telephony can afford only error-resilience techniques, whereas high-delay or high bandwidth applications can afford complete error recovery using either forward error correction or retransmission-based techniques.

It is worth noting here that although applications in all three classes play out multimedia in real time, the phrase "real-time communication" is commonly used only for the first application, that is, audio and video telephony, conferencing, and gaming, whereas the phrase "streaming" is often associated only with the latter two applications.

1.1.4 Online Versus Off-Line Encoding

Another essential difference between multimedia communication applications is whether the content is encoded online, as in the case of real-time communication or live broadcast applications, or is encoded off-line, as in the case of streaming media on demand. The advantage of online encoding is that the communication channel can be monitored and the source and channel coding strategies can be adapted correspondingly. For instance, the receiver can inform the transmitter of the information that is lost and the encoder can adjust correspondingly. The advantage of off-line encoding is that the content can be exhaustively analyzed and the encoding can be optimized (possibly in nonreal time over several passes of the data) for efficient transmission.

1.1.5 Receiver Device Characteristics

The constraints of the receiver devices on which the various applications are consumed by the end user also have an important impact on multimedia communication. In particular, the available storage, power, and computational capabilities of the receiving device need to be explicitly considered when designing complete multimedia communication solutions. For instance, the design of multimedia compression, scheduling, and error protection algorithms at the receiver should explicitly consider the ability of the receiver to cope with packet loss. Also, receiver-driven streaming applications can enable the end device to proactively decide which parts of the compressed bit streams should be transmitted depending on the display size and other factors.

1.1.6 Unicast, Multicast, and Broadcast

Multimedia communication can be classified into one of three different categories: unicast, multicast, and broadcast, depending on the relationship between the number of senders and receivers. Unicast transmission connects one sender to one receiver. Examples of such applications include downloading, streaming media on demand, and point-to-point telephony. A main advantage of unicast is that a back channel can be established between the receiver and the sender. When such a back channel exists, the receiver can provide feedback to the sender about experienced channel conditions, end-user requirements, end-device characteristics, and so on, which can be used accordingly to adapt compression, error protection, and other transmission strategies.

Multicast transmission connects the sender to multiple receivers that have elected to participate in the multicast session, over IP multicast or application level multicast. Multicast is more efficient than multiple unicasts in terms of network resource utilization and server complexity. However, a disadvantage of multicast compared to unicast is that the sender cannot target its transmission toward a specific receiver.

Broadcast transmission connects a sender to all receivers that it can reach through the network. An example is broadcast over a wireless link or a shared Ethernet link. As in multicast, the communication channel may be different for different receivers. In this book, when we refer to the live broadcast application, we are usually talking about a solution in which a live signal is actually multicast over the network.

1.1.7 Metrics for Quantifying Performance

Unlike conventional communication applications, multimedia communication applications cannot be simply evaluated in terms of the achieved throughput, packet loss rates, or bit error rates, as these applications are delay sensitive and not all the various transmitted bits are "created equal," that is, have the same importance. Instead, multimedia performance needs to be quantified in terms of metrics such as the perceived quality or objective metrics such as the Peak Signal-to-Noise Ratio (PSNR) between transmitted and received media data. Hence, the importance of each bit or packet of multimedia data depends on its delay requirements (i.e., when it needs to be available at the receiver side) and impact on the resulting PSNR. These new evaluation criteria fundamentally change the design principles for multimedia communication systems compared to communication systems for traditional delay-insensitive, loss-intolerant applications.

1.2 ORGANIZATION OF THE BOOK

This book aims at providing an in-depth understanding of the theoretical foundations, key design principles, algorithms, and existing standards for the aforementioned multimedia networking and communication scenarios. The book is divided into five major parts.

The first part of the book discusses how multimedia data can be efficiently compressed to enable optimized transmission over the Internet and wireless networks. Unlike traditional compression techniques such as MPEG-2, which were designed solely for storage (e.g., on DVD disks) or transmission over error-free networks with relatively large and guaranteed bandwidth, compression schemes that enable efficient multimedia communication over the Internet and wireless networks need to have the ability to cope with different channel conditions, characterized by different bit error rates, packet loss rates, access bandwidths, or time-varying available bandwidths. Chapter 2 discusses error-resilient techniques for video transmission over such error-prone networks, while Chapter 3 presents algorithms and solutions for error-resilient audio transmission. To cope with the changes in bandwidth, Chapter 4 provides a thorough analysis of the various mechanisms for bandwidth adaptation, as the network often offers heterogeneous, time-varying channel conditions. To effectively cope with adaptive streaming applications or multicasting applications, where a variety of receivers would like to simultaneously access the same multimedia content, Chapter 5 introduces existing and emerging scalable video coding algorithms, while Chapter 6 discusses scalable audio coding.

The second part of the book focuses on efficient solutions for bit stream transmission over IP networks. Chapter 7 introduces the fundamentals of channel protection needed to insulate bit streams from the error-prone nature of the channels over which they are transmitted. Chapter 8 discusses how to effectively model and characterize the complex communication channels within networks such as the Internet. Having an accurate model of the channel becomes paramount when finding an efficient trade-off between the bit rates allocated to source and channel