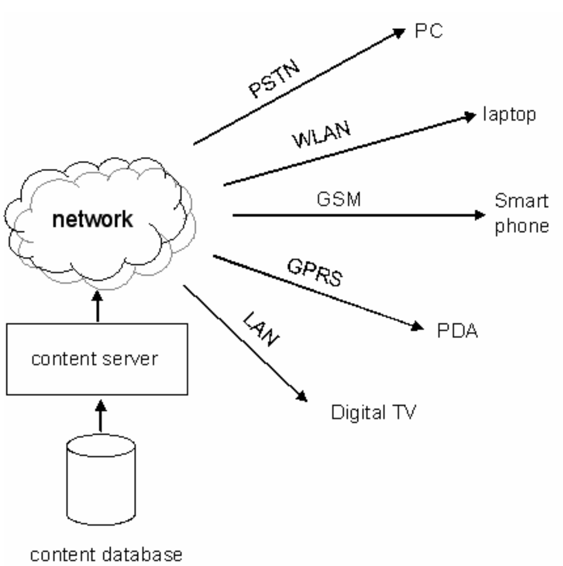
Universal Multimedia Access

Andrea Cavallaro
Queen Mary, University of London, UK

INTRODUCTION

The diffusion of network appliances such as cellular phones, personal digital assistants (PDAs), and handheld computers creates a new challenge for multimedia content delivery: how to adapt the media transmission to various device capabilities, network characteristics, and user preferences. Each device is characterized by certain display capabilities and processing power. Moreover, such appliances are connected through different types of networks with diverse bandwidths. Finally, users with different preferences access the same multimedia content. To cope with the challenge of delivering content to such a variety of conditions while maximizing user satisfaction, multimedia content needs to be adapted to the needs of the specific application, to the capabilities of the connected terminal and network, and to the preferences of the user (Mohan, Smith, & Li, 1999a; Van Beek, Smith, Ebrahimi, Suzuki, & Askelof, 2003). This adaptation enabling seamless access to multimedia content anywhere and anytime is known as universal multimedia access (UMA). The UMA framework is depicted in Figure 1. Three main strategies for adaptive multimedia content delivery have been proposed, namely, the info pyramid, scalable coding, and transcoding. These strategies, emerging trends in UMA and standardization activities, are discussed in the following sections.

Figure 1. Universal multimedia-access framework (PSTN: Public Switched Telephone Network; WLAN: Wireless LAN; LAN: Local Area Network; GSM: Global System for Mobile Communications; GPRS: General Packet Radio Service)



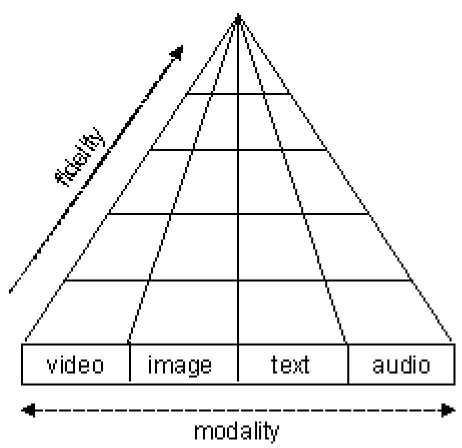
INFO PYRAMID

Traditional solutions to multimedia adaptation encode and store multimedia content in a variety of modalities and formats that are expected to fit possible terminals and networks (Li, Mohan, & Smith, 1998). The most adequate version is then selected for delivery according to the network and hardware characteristics of the specific appliance. The advantage of this approach is speed of access because the content is already available and does not need to undergo any transformations. On the other hand, the limitation of this approach is the difficulty of generating a distinct content version for each profile of capabilities for the large variety of terminals and networks currently available. A general framework for managing and manipulating media objects is the info pyramid. The info pyramid manages different versions, or variations, of media objects with different modalities (e.g., video, image, text, and audio) and fidelities (summarized, compressed, and scaled variations). Moreover, it defines methods for manipulating, translating, transcoding, and generating the content (Smith, Mohan, & Li, 1999b). When a client device requests a multimedia document, the server selects and delivers the most appropriate variation. The selection is based on network characteristics and terminal capabilities, such as display size, frame rate, color depth, and storage capacity. The info pyramid of a media object is defined as a collection of the different variations of that media object, as shown in Figure 2. A content value score is then associated to each media object. The value score is assigned manually or based on some automatic measure, such as the entropy. Finally, the most appropriate media object is selected by maximizing the total content value for a set of device and/or network constraints. Utility-based frameworks are generally developed for the selection mechanism. In Mohan, Smith, and Li, (1999), the rate-distortion framework is generalized to a value-resource framework by treating different variations of a content item as different compressions, and different client resources as different bit rates. With the info pyramid approach, higher quality or higher resolution bit streams repeat the information already contained in lower quality or resolution streams. Then additional information is added to manage the streams. For these reasons, the info pyramid is not efficient. To overcome this problem, the redundancy should be removed by coding multiple fidelity levels into a single stream, as described in the next section.

SCALABLE CODING

As opposed to the info pyramid, scalable coding processes multimedia content only once. Lower qualities or lower resolutions of the same content are then obtained by truncating certain layers or bits from the original stream (Wang, Osterman, & Zhang, 2002). In the case of video, basic modes of scalability include quality scalability, spatial scalability, temporal scalability, and frequency scalability. These basic scalability schemes can be combined to reach fine-granularity scalability, such as in MPEG-4 FGS (Fine Granularity Scalability) (Motion Pictures Expert Group; Li, 2001). Quality or SNR (Signal-to-Noise Radio) scalability is defined as the representation of a video sequence with varying accuracies in the color patterns. This is typically obtained by quantizing the color values with increasingly finer quantization step sizes, as shown in Figure 3. Spatial scalability is the representation of the same video in varying spatial resolutions. Corresponding layered bit streams are usually produced by computing a multiresolution decomposition of the original image. Next, the lowest resolution image is coded directly to produce a first layer. For each successive layer, the image from the previous layer is first interpolated to the new resolution, and then the error between the original and the interpolated image is encoded. Temporal scalability is the representation of the same video at varying temporal resolutions or frame rates. The procedure for producing temporally layered bit streams is similar to the procedure used for spatial scalability, but temporal resampling is used instead of spatial resampling. Frequency scalability includes different frequency components in each layer, with the base layer containing low-frequency components and the other layers containing increasingly high-frequency components. Such decomposition can be achieved via frequency transforms like the DCT (Discrete Cosine Transform) or wavelet transforms.

Figure 2. Multimodal representation of a media object as a collection of different variations of the same object in the info-pyramid approach



TRANSCODING

Transcoding is the process of converting a compressed multimedia signal into another compressed signal with different properties (Vetro, Christopoulos, & Sun, 2003). Unlike the info pyramid and scalable coding, transcoding can operate according to the current usage environment on the fly without requiring a priori knowledge of terminal and network capabilities. Early solutions to transcoding determined the output format based on network and appliance constraints only, independent of the se-

Related Content