When can we unplug the radio and telephone?

Henning Schulzrinne
GMD Fokus, Berlin
hgs@fokus.gmd.de

Abstract

Despite high-performance workstations, special-purpose communication devices like telephones and radio still dominate. Reasons include inadequate system and network support, both in terms of performance and the ability to write applications without specialized signal processing knowledge. The interaction of applications with the network will depend largely on the future tariff structures rather than protocol issues.

This paper briefly summarizes some open issues that stand in the way of having a true convergence of computers and audio/video communications. Most of these problems are actually engineering, rather than research, issues.

1 Audio

The availability of graphical user interface toolkits has significantly accelerated computer deployment and new applications. The same reasoning has led to a number of efforts to build the equivalent of Xlib for audio. Experience in porting the NeVoT audio tool to different platforms has unfortunately shown that the easiest to work with are those sticking most closely to the traditional Unix file model, allowing integration into the X event handler mechanism or simple blocking read/write. Some toolkits force the use of timers to continuously read audio, causing overhead and synchronization problems.

An improved architecture for an audio server distinguishes a real-time path, where local and remote audio applications communicate via a non-flow-controlled, unreliable network association (like RTP over UDP) instead of TCP, from a path used to simply download audio clips for later play. A control path would be used to set the local audio encodings, initiate playback of audio clips and carry events such as audio activity indication or VU meter values.

Other desirable changes for audio support include:

- interfaces with independent input/output and setable DMA sizes, possibly even built-in echo cancellation;
- standard system libraries for DSP tasks like AGC, rate conversion, silence detection;
- CPUs with efficient support for DSP operations (fast multiply-adds);
- predictable delays and timestamps so that input and output audio samples can be correlated in time;

NOSSDAV'95 (5th International Workshop on Network and Operating System Support for Digital Audio and Video); Durham, New Hampshire, April 18–21, 1995 (pp. –)

2 Video

Most current systems force a copy of 24-bit uncompressed video across the system bus to the display adapter. It should be possible to integrate MPEG decoders or at least YUV-to-RGB converters and dithering into video display systems. For smooth motion, asynchronous framegrabbers must be replaced by interfaces that signal the availability of new video frames or at least include timing information with the digitized image. Operating systems must support bounds on the process scheduling jitter, so that the period between invocations never exceeds the buffering depth of the video adapter.

3 Network Support

Standard local area networks like Ethernet or token ring can serve audio and video reasonably well, particulary in their switched versions. However, multicast should enjoy the same learning-bridge filtering as unicast.

In wide-area networks, support for some form of resource reservation (with the attendant charging mechanisms) or very low network load is necessary for acceptable media quality. Despite claims, ATM still falls short in this regard. The current ATM UNI signaling requires adding each user individually to a point-to-multipoint connection, which is cumbersome and scales poorly; future versions may allow receiver-initiated joins.

With the appropriate financial inducement, many users would be willing to reduce network usage during periods of congestion. Current networks provide only crude information (packet loss) on network loading; further work is needed to determine whether ABR-like feedback mechanisms are useful, particularly in the case of multicast. Also, multicast combined with the increasing range of network speeds calls for scalable media codings, with RSVP-style filters or RTP-style translators at low-speed network access points. Advance scheduling, desirable for important applications, changes the reservation model in that newly arriving reservations might be given a time-limited rather than open-ended commitment.

The actual use of reservations will largely depend on the tariff model. For example, peak-rate reservations avoids the difficulty of characterizing VBR video and may be attractive if there is a rebate for unused bandwidth.

It is not yet clear whether *near-real-time* applications like video-on-demand or distance learning are best served by TCP-like protocols, with the danger of starving the receiver but increased network utilization and guaranteed reliability, or some hybrid protocols which offer some retransmission capability and some rate flexibility. The use of packet FEC in real networks bears further scrutiny.

Virtual reality combines the stringent delay requirements of interactive multimedia (to achieve rapid response to user actions) with the high quality requirements of video-on-demand.