How the Internet will Replace Broadcasting
by Edmund X. DeJesus

Five-hundred cable channels. Scheduled pay-per-view events. Interactive TV on demand. Who cares?

Try an <u>unlimited</u> number of channels. Whatever you want, when ever you want it. From anywhere on the planet. For free.

Internet broadcasting will bring real-time audio and video -- radio and TV -- to modest desktop machines over ordinary phone lines. Not download-for-20-minutes-and-play-later clips, but audio and video streaming through the wires in real time.

Internet broadcasting is overcoming technical obstacles like the narrow bandwidth of phone lines, the limits of compressing multimedia data, and the vagaries of Internet packet transmission.

The selection of multimedia available over the Internet is surprisingly varied, considering that some of the technologies that support it are only about a year old. You can listen to live and recorded news and sports from huge networks like ABC, CBS, ESPN, and NPR. You can watch live news video feeds from NBC. You can tap into music from major recording companies and fledgling bands. And, as with all things Internet, you can find home-grown, impossible-to-categorize sights and sounds with all the immediacy of real time.

Strike Up the Bandwidth

Pumping full-motion video over the Internet is not a fun task. Do the math. A 1024- by 768-pixel display (good for a monitor, lame for a movie) with three colors at 8 bits apiece, running at 30 frames per second, means at least 566,000 Kbps hurtling down the wire. Real-time audio is simple by comparison. CD-quality sound generally consists of 16-bit samples, 44,100 samples per second, for a mere 706 Kbps. Digitized phone-quality speech is only 64 Kbps (8-bit samples, 8000 samples per second).

Houston, we've got a problem: Even the best plain old telephone system (POTS) can handle only about 100 Kbps of data. Worse, today's modems top out at 28.8 Kbps, and tomorrow's products don't look much better. Current and next-generation modems are already hitting a ceiling in the 30- to 40-Kbps range, says Nicole Toomey Davis, product line manager for modem vendor Megahertz (a subsidiary of U.S. Robotics). Clearly, you aren't going to be putting out many fires with this soda straw.

Fortunately, there are alternatives to ordinary phone service. Unfortunately, access to those alternatives varies. So does their cost. Everybody has analog POTS. The price and availability are certainly right, but, as the math shows, you can't fit much down the narrow pipeline that POTS provides. Signaling (status information about the modem link, for example) may be in-band,

chewing up yet more bandwidth.

A step up from POTS is ISDN. While many people still think of ISDN as technology I Still Don't Need, it has a lot to offer the Internet broadcaster and broadcastee. ISDN is an international telecommunications standard for transmitting voice, video, and data over digital lines. Circuit-switched bearer channels (B channels) carry voice and data at nominal rates of 64 Kbps. (Actual rates in the United States are 56 Kbps due to the way older equipment handles switching.) A separate data channel (D channel) of either 16 or 64 Kbps carries control signals that would, with POTS, take up in-band bandwidth. (For actual phone service, the dedicated D channel carries information relating to special features like call forwarding or call waiting. This dedicated D channel also enables ISDN modems to connect to each other more quickly than analog modems.)

There are several varieties of ISDN service. Basic Rate Interface (BRI) has two B channels and one D channel (written 2B+D in ISDN shorthand), for a theoretical total of 144 Kbps and an actual rate of 128 Kbps. That's nearly five times the bandwidth of a 28.8-Kbps modem. In North America, the ISDN Primary Rate Interface (PRI) has 23 B channels and one beefed-up 64-Kbps D channel (23B+D) for a total of 1544 Kbps. This is equivalent to a T-1 line, over 50 times the bandwidth of a 14.4 modem, and well into the CD sound range (but still nowhere near the capacity needed for raw video). In Europe, the ISDN PRI has 30 B channels and one D channel (30B+D), for a breathtaking 2048 Kbps -- equivalent to the European E-1 line service.

T-1 is a fast (1544 Kbps) but point-to-point mechanism: Your T-1 box talks only to one specific T-1 box somewhere in the world, and it is always talking to that box. With ISDN, you can dial up any other ISDN site just as with a phone: It's a so-called "cloud" architecture, just like the phone system. You dial into the phone company cloud, and someone somewhere somehow gets your ISDN call. Other T carrier systems include T-1C (3152 Kbps), T-2 (6312 Kbps), T-3 (44,736 Kbps), and T-4 (274,176 Kbps). As with ordinary phone service, T-1 transmits 8000 frames per second; the difference is that a T-1 frame is 193 bits long, enough for 24 8-bit samples and one synchronization bit. This is serious firehose territory.

The big kahuna of bandwidth is asynchronous transfer mode (ATM), a cell-switching technology. ATM rates begin at 1544 Kbps, spiral up through 25,000 Kbps and 155,000 Kbps, winding up at 622,000 Kbps today and will maybe go beyond that tomorrow. ATM uses small, fixed-length, 53-byte cells (kind of like packets). The 5-byte header contains a CRC code for error control, address information, and priority control codes. The lower 48 bytes contain the data. Since the cells are of fixed length, switches can be very fast.

While ATM may sound like broadcast bandwidth nirvana, it has some problems. First, ATM is not universally available. Second, the standards are not yet clear. Third, it's expensive. Many analysts predict that the next three to five years will see changes that fix these problems.

But wait. Phone companies aren't the only ones with firehoses to every home and office. Cable companies want you to access the Internet through their cables (essentially bypassing their own programming). To do this, you need 1) a cable company that is actually supplying Internet access

and 2) a cable modem. AT&T, Intel, Hybrid Networks, Hewlett-Packard, LANcity, Motorola, and Zenith Data Systems sell cable modems for several hundred bucks apiece. Speed reportedly is at least 500 Kbps, and there are claims of up to 30,000 Kbps. Clearly, cable would be a very convenient solution for people with access to such a provider, but you'd have to buy the cable modem and appropriate software, and pay the additional cable fee.

Compression: The Big Squeeze

Raw bandwidth is one thing. What you do to optimize it is another. Recently, you could have tuned in to Scott Cook, CEO of Intuit, as he gave a speech in New York. The broadcast was live, in somewhat-real time, and would have come to your computer using Xing Technology's StreamWorks video and audio software. No ATM. No T-1. No ISDN. Just a vanilla 14.4-Kbps modem.

The image with this approach is about 3 by 4 inches, rather Impressionistic in its graininess, and changes every four or five seconds. Sound is AM-radio quality. But, like a dog talking with a lisp, it's impressive, even if imperfect.

How can sound and video pour through such a pitiful spigot as this? The answer begins with compression.

Data compression is all over computers, of course. You routinely ZIP or Stuff-It unruly files to save room. Your hard disk may well be a compressed drive. Your modem probably compresses data before transmitting it. What makes Xing's   and VocalTec's Internet Wave and Progressive Networks' RealAudio and VDOnet's VDOlive   compression and decompression remarkable is that it happens in near-real time. Instead of the typical download-and-run method of getting multimedia data from a network, your computer opens a connection to the server and starts decompressing and playing the data it's off-loading over the wire.

But the constraints on this compression/decompression are heavy. Since what we want is watchable video and listenable music, the compression must deliver fairly high-detail results. Very lossy algorithms lose too much, while lossless algorithms like Lempel-Zev don't compress enough. Both the compression and decompression algorithms must be very fast to permit live broadcasts, which eliminates super-crunch algorithms that take too long.

And since compressed packets will travel via Internet, with no guarantee of arrival, later packets can't depend on previous packets. This excludes many efficient algorithms that are based on lookup tables of symbols and their expansions.

In fact, variable compression rates based on access method -- more for a 14.4-Kbps modem and less for an ISDN line -- make the most sense. For example, RealAudio's claimed compression rates range from 8:1 to 22:1 depending on the access method.

Don't think that compression by your modem is going to bail you out. As Mike Peterson, product manager for Megahertz, points out, "You can't compress already-compressed data very

much: possibly by 50 percent but not more."

And because you don't want users to have to buy any new hardware, you have to choose a software-only solution that will run on standard desktop platforms. No cheating with fancy dedicated decompression chips.

With these kinds of constraints, it's no wonder that vendors tend to depend on prior work. VocalTec, for example, has built on its experience with lossy algorithms gained from its Internet Phone (I-Phone) product as well as previous chip products. Human speech has all kinds of features that permit efficient lossy-but-satisfactory compression. For instance, we may pause slightly between words: That's dead air, and no need to record it. We speak in a narrow range of frequencies, and the full dynamic range, from whisper to scream, can be reduced without much loss of understanding. And usually only one voice is speaking at a time.

Music is much tougher. There may be no pauses (rests) at all in a musical piece. Multiple sounds are the norm, not the exception. The frequency (pitch) range and dynamic range of instruments in a single piece -- indeed, at a single moment -- can be wide. There can be a soft piccolo and a screeching guitar at the same time.

The result is that spoken words -- a news report, sports play-by-play, lecture, sermon, interview, and so forth -- generally sound pretty good on on-line "radio" broadcasts. It's the music that sounds like a warbly AM station.

MPEG and Beyond

Video compression's standard is MPEG encoding. One nice thing about video is that often, large portions of a scene are unchanged from frame to frame. A succeeding similar frame can be stored with a symbol meaning "ditto" or "ditto except this arm moved." Thus, MPEG compression is about three times more compact than a sequence of, say, JPEG-compressed frames. The more similar the frames, the more compact the result. A fixed-mount camera view of a lecture will compact smaller than an action movie would.

Since you can't depend on the arrival of a previous packet before decompressing the current one, however, MPEG isn't ideal for Internet transmission. MPEG uses discrete cosine transform (DCT) algorithms that, like fast Fourier transforms, essentially decompose data into sets of wave frequencies. The compression process retains only certain principal frequencies and discards less important ones. You lose some detail in the process. (Then again, you lose some detail whenever you record reality.) The question is whether the level of detail retained is sufficient for your purpose. Merely seeing another person's face updated occasionally may suffice for a video-conference or a lecture, while extreme detail is preferable when watching movies or deep technical information. MPEG-1 supports 320 pixels by 240 pixels of three-color, with 8 bits per color at 30 frames per second, and CD-quality sound. (Some vendors use MPEG-1 to compress video for CD-ROMs.) MPEG-2 is an emerging standard intended to reproduce full-screen, broadcast-quality video and sound.

MPEG-1 generally requires more processing to encode video than to decode it, making live video more difficult to compress compactly. The maximum MPEG-1 compression is about 200:1, but 50:1 is more typical. Thus, maximum MPEGed video might require bandwidth of under 4000 Kbps to be useful. Drop to black-and-white and you're in the high ISDN, low T-1 ballpark.

MPEG-2 is even more TV-oriented than MPEG-1. MPEG-2 knows how TV "frames" interlace, for example. Picture quality is better with MPEG-2, also. But the bandwidth problem makes good-enough MPEG-1 preferable to MPEG-2 as a distribution compression scheme.

New compression tools, such as wavelets or fractals, will find use in Internet broadcasting. (See **News & Views**, December '95, page 34.) Microsoft and Intel are reportedly using wavelet technology in their respective "Blackbird" and Indeo products. Some research projects have produced nearly 500:1 compression of video, but not in a commercial product -- yet. Since compression techniques continue to change and improve, it is important to retain the flexibility afforded by software-only solutions. It is also important to be able to swap the compression algorithm in browsers and other software, should a hot new one appear.

Safe and Sound

The main application of real-time compression on the Internet is speech. In hearing normal speech, it doesn't bother us much if we occasionally don't catch every syllable. It may be mildly annoying, but we generally interpolate what the person probably said. It's different with music. Missed notes disrupt enjoyment. Since the Internet doesn't guarantee that packets containing music are going to arrive in time to play (or even that the packets are going to arrive at all), music is a problem on the Net. How can the audio provider ensure that customers aren't getting stuck listening to portions out of sequence or, worse, to dead air?

The basic problem has to do with the Internet as a delivery medium. Yes, it's a fast, scaleable, packet-switched network. But it is not designed to handle isochronous (continuous time-based) information. The Transmission Control Protocol (TCP) endeavors to guarantee packet delivery, but delays in delivery may occur when a server retransmits a packet to a client, or when waiting for the client to acknowledge receipt. On the other hand, the User Datagram Protocol (UDP) does not take the precautions TCP does to guarantee delivery. UDP ships out a stream of packets with as little delay as possible, risking the occasional lost packet. Neither guarantees throughput rates or latency periods. Which protocol do you choose?

VocalTec's Internet Wave uses TCP. Packets should all show up, but some may be late. One way to deal with late packets is with a sufficiently large buffer -- VocalTec uses a predictive cache. If enough music has accumulated in the buffer, the late packet may show up before its turn to play. Internet Wave then inserts it into the sequence and the listener is none the wiser. If there are more extensive or systemic delays in packet delivery, the sound breaks up more seriously. It's like the sound of an AM car radio when you drive under a bridge.

Progressive Networks' RealAudio uses UDP. Packets arrive quickly, but the delivery can be less reliable. To compensate, RealAudio interleaves the information. It takes about 3 seconds of

sound and divides it into 144 bundles of 20 milliseconds each. Then it distributes the bundles among 12 packets. The first bundle goes in the first packet, the second in the second, proceeding until the thirteenth bundle goes back in the first packet, the fourteenth in the second, and so on. Each packet thus consists of twelve 20-ms bundles, plus an information header. With this scheme, if a packet vanishes, you don't lose a full quarter-second of sound; instead you lose 20 milliseconds every quarter of a second for about 3 seconds (or until the errant packet shows up). Such a disruption is minor compared to a quarter-second gap.

There's another difficulty with using Hypertext Transport Protocol (HTTP) as a transport for audio: The Web isn't inherently bidirectional. When viewing a video or listening to music, you may wish to rewind, fast-forward, or resume playing, just as you do with a VCR. Problem is, HTTP does not support return commands.

There are several ways around this. VocalTec sticks with HTTP, implementing user control with a Common Gateway Interface (CGI) file on the HTTP server. With CGI, a program or script can run on the server and return output to the client. (A common example of this is the result of a search you perform on Yahoo or Lycos that returns an HTML-format page for your perusal.) In this case, user input triggers a program on the server that carries out the request.

The alternative is to not use HTTP. For example, RealAudio has its own protocol and its own separate nonWeb server to field requests for RealAudio transmissions. This separate server can actually be located on the same physical machine as the host Web server. The RealAudio protocol supports bidirectional communication between client and server. Thus, to fulfill a request from a client to the Web server, the Web server triggers a request to the RealAudio server, which then returns the material to the client. One can imagine third-party RealAudio servers dedicated to fielding requests from Web servers handling clients. The upside of this is a separation of logical server functions, and the employment of protocols especially suited to the situation, namely, transmitting sound. The downside is it means another server to maintain and a proprietary protocol to understand.

I'll Be Your Server This Millisecond
So, what would you put into a server that's supposed to be feeding a gush of video onto the Internet? A fistful of screaming RISC chips? Enough memory to build a dozen desktops? Not bad for a start. But the real secrets of designing servers to pump video are in some unglamorous places: internal bandwidth, hard drive performance, and the operating system.

Sun Microsystems, for example, has an operating system that's fine-tuned for on-line broadcasting. "We use a special version of Solaris, optimized for real-time I/O," explains Steven Kleiman, chief architect at Sun. This special multithreaded OS has a streaming driver (called a "bit pump") in the kernel. In most systems, I/O is something the operating system can't predict and requires a complex structure of interrupts to handle. In Sun's MediaCenter family of servers, the OS treats I/O as a regularly recurring process. The operating system schedules the I/O to get the most bandwidth from all the subsystems, including the hard drives. Because this is not a general-purpose OS, but one optimized for high-bandwidth I/O, it can get out of the way of a lot

of the processing.

"In Sun's experience," says Anne Schowe, general manager of Sun's interactive systems group, "serving video stresses the internal bandwidth more than the CPUs." Sure, Sun offers multiple CPUs (the 1000E model has four SuperSparc+ chips), but there is minimal central processing going on. The main activity is disk-to-output, as much and as fast as possible.

How to speed disk access? The main answer lies in RAID technology. Sun's RAID level 4 system can handle multiple I/O requests from an appropriately enabled operating system (like the tweaked Solaris). With multiple Fast SCSI-2 2.1-GB drives streaming data simultaneously, the video flows nicely. The top-of-the-line 1000E moves the bits at 400,000 Kbps (equivalent to about 100 MPEG-2 streams or 270 MPEG-1 streams).

Becoming an audio broadcaster on the Internet requires far less sophisticated, and less costly, hardware. Audio-server software, like Internet Wave or RealAudio, will run on a high-end 486 or Pentium PC.

What's Out There?
This ability to broadcast with modest equipment has astonishing implications. How much lucre would it cost to have your own local radio station? That's why there aren't that many. But if the entry requirements for global Internet audio broadcasting are in the mere thousands of dollars, we may see an explosion of audio broadcasting sites. With more than 500,000 Web sites up now (a figure increasing at a staggering rate), even if only one-hundredth provide some real-time audio capability, we're talking about 5000 new broadcasters whose message is audible anywhere on the planet at any time. That's significant.

What kinds of broadcasting are we likely to see? Currently, the best uses for Internet audio involve speech, and there are plenty of possibilities. News organizations like ABC, NBC, and NPR are already posting live news broadcasts. They and others are also providing interviews, weather, and editorial content. Sports play-by-play makes a lot of sense. Whereas current sports broadcasting is usually local, the theory being that no one in California wants to follow the Boston Red Sox, Internet sports broadcasting would allow fans to tune in from wherever they are. Yes, you could be in Singapore and listen to the Sox blow the pennant.

Educational and other social purposes may become important. Schools can transmit lectures. Political speeches go out live. Radio drama. Language lessons. Anything spoken can be put on the Web.

Although not the highest-quality music distribution medium, Internet broadcasting of music definitely has its place. Bands who used to make demo tapes can now have Web pages where listeners can sample their sounds. Radio stations are already simulcasting some of their programming on the Internet. Record companies could post samples for fans to check out.

Due to the current low frame rates of Internet video -- and that will improve only as

compression and bandwidth do -- you're not likely to be watching _The Brady Bunch_ on a Web site anytime soon (thank goodness). The best application areas are those where nonmoving pictures will work. A university lecture. Remote monitoring.

One surprising possibility is videoconferencing for the masses. White Pine Software's CUSeeMe can hook up multiple video-camera-equipped locations using 14.4-Kbps modems. The users see multiple windows of everyone in the conference, with images updating every few seconds. Sure beats flying.

Businesses can use existing video capabilities for some interesting additional purposes. Rather than distributing training videos to a hundred locations, you can post the video on a server where employees can log in at their convenience; this would be especially valuable for sales staff who need to familiarize themselves with a new product line rapidly. But customers could also browse on-line animated catalogs that deliver a narrated demonstration. Anything that the much-maligned slide presentation can do, existing video capabilities can do on the Internet, in real time, at the user's convenience and control.

Unleashing audio and video on the Internet may degrade be the whole Net's performance may degrade as a result. No one really knows the long-term effects, but it will hasten the need to upgrade the Internet infrastructure. Commercial users of Internet broadcasting for internal purposes may want to use one of the service enablers like Concentric Network. These private Internet-accessible services can provide guaranteed levels of bandwidth and security that the Internet cannot. That will be attractive to businesses who want the convenience of putting company information on-line but don't want people outside the company accessing it.

Tune In...When?
"No future." "Who wants that?" "Commercially and financially impossible." No, these aren't comments about the prospects for Internet broadcasting. They are the opinions of Lord Kelvin, Harry Warner (Warner Brothers), and Lee deForest on the future of radio, motion pictures with sound, and television, respectively. Sure, right now Internet television is a jerky black-and-white postage stamp. But commercial TV began with a rotating plastic statue of Felix the Cat. Could you extrapolate from that minimalistic presentation to current television's dominant cultural and commercial influence?

You may have missed out on Samuel Morse tapping out, "What hath God wrought?" and Alexander Graham Bell yelling, "Mr. Watson. Come here. I need you." But you are present at the dawn of the Internet Broadcasting Age. Keep your browsers tuned.

- Edmund X. DeJesus is a senior technical editor at BYTE. You can send E-mail to him at edejesus@bix.com.