AES47

The AES has released AES47, a new standard for passing uncompressed professional audio over high-speed networks. Typically using ATM 25Mbs or 155Mbs interfaces and conventional Cat 5 structured wiring or optical fibre, AES47 can interoperate with other types of data to provide a complete communications package for the most challenging audio and video tasks.

JOHN GRANT, NINE TILES NETWORKS

AUDIO ON TRADITIONAL cables provides a service, but the cables themselves are inflexible, bulky and expensive. Networks, on the other hand, are simpler to install, offer large-scale routing inherently, and can be engineered to be fail-safe. So if networks are so great, why has it been so hard to use them before for audio?

Almost all computer networks in recent memory are based on Ethernet Local Area Networks and the Internet. Not so long ago, networking was a serious tool used by serious corporates to connect terminals to mainframe computers. Typically, each computer manufacturer had its own network scheme and compatibility was not considered a driving issue. With the rise of personal computers, the pressure towards networks with standardised interfaces became irresistible. One standardised networking system, IEEE 802.3. (more commonly called Ethernet), was just one contender but in time succeeded in dominating the office data market and established the largest installed base worldwide. Although these standardised LANs were still mainly independent islands, some were beginning to interconnect.

The explosive growth of the Internet accelerated everything very rapidly. Suddenly, it was possible for all computers to be connected to all other computers in the world. This was possible because the Internet used available technology to appeal to the greatest community of potential users, which were initially military and academic establishments, then business and general users. They almost all used Ethernet LANs.

The early Internet applications were limited not so much by the available computer hardware or software, but by the low speed of the interconnections. While politicians were making brave statements about the 'Internet super-highway', the reality was throttled to the speed of a 14.4kbs modem. This was fine for transferring text-based documents and simple Web interactivity but was a far cry from the audio visual promises that were being made. The Internet protocol was sufficiently good to get the Internet going as a useful tool, the ball was rolling, and the real killer application was email.

Now that we have reasonably fast connections, the 'super-highway' is closer to reality, but the network protocol is under increasing strain. We are now being offered audio and video streams but, to overcome the limitations of Internet delivery, they are structured in very different ways to professional streams and they have quite different criteria for success.

Whether compressed or not, audio needs to arrive in a continuous and unbroken stream. The real-world bandwidth restrictions of the Internet make this practical for heavily data-compressed audio, but very difficult for the uncompressed audio data you would want in your production. In addition, live audio needs to arrive in a continuous stream with no appreciable delay.

The Internet packs data into conventional Ethernet packets, each with header data to indicate its final delivery address. The useful packet content, or 'payload', is usually kept large compared to the fixed size header data to keep the administrative overhead down and to use network bandwidth more efficiently. Each of these packets makes its way from router to router across the Internet to its intended destination. At each router, the packet's address information is read and checked against a list of locally-available options. The packet is then retransmitted along the most appropriate network segment chosen from the options available, taking into account the needs of other traffic at that instant. If a packet takes too long

Samples being packed into cell

Cells in transit

Unpacking samples from cell

to reach its destination, or is undeliverable for any reason, it will simply be dropped and the two computers exchanging that data will need to arrange for another packet to be sent or accept a hole in the received data.

This is the 'connectionless' principle that underpins the Internet. The good part is that it doesn't matter if routers or network segments are damaged, the packet can always get through by another route automatically. But the large size of Ethernet packets coupled with the lack of a firm connection is also the root of our problem for live audio. They all contribute to an overall delay, which rapidly becomes unacceptable.

Look at Internet radio for a demonstration of what I mean. Internet radio uses high compression coding and it takes significant time to encode at the broadcaster and to decode at your receiver. So, understanding the relevant coding technologies, you would expect a delay in the order of 100 milliseconds or so to be incurred by the coding process. To find out what actually happens try the following.

Tune an analogue AM or FM radio to a convenient radio channel, go on-line and find the Internet version of the same radio channel, and compare the timing difference. The delay between an announcement on the analogue radio and the Internet equivalent is not small; it's typically many seconds and can run to minutes. Imagine trying to hold a two-way interview with communications delays like this. A stream, maybe, but a 'live' stream? Perhaps not.

The delay between an audio sample being received at the transmitting network interface and that same audio sample being output by the receiving network interface has three components: packetisation delay; transit delay; and buffering delay.

Packetisation delay is the time between the first sample being received and the packet being ready for transmission. In the first diagram, it's the time between the first person getting on the bus and the bus being ready to leave.

If samples arrive at regular intervals and n samples are packed in a packet, then the packetisation delay will reasonably be n sample times.

Transit delay is the time from the packet being ready for transmission until it is ready to be unpacked at the receiving end (or from the bus being ready to leave until the bus arrives at its destination). This is composed of the time from starting to arrive at each switching or routing element in the network until it can start to be transmitted, time waiting to be transmitted on each link, and the propagation delay along the link (time queuing at road junctions etc, waiting for traffic lights, and actually moving).

The length of the link and the speed of the signal along it govern the propagation delay, which is about 1m per 5 nanoseconds or 200km per millisecond. The length of the cable may be considerably more than the straight-line distance between its ends. Buffering delay is the time between the packet arriving at the destination and the first sample being output.

The receiving device needs to output samples at regular intervals, and needs to be sure that each sample will have arrived by the time it is needed. It therefore needs to keep some data in hand in case a packet arrives late (there needs to be either a queue of buses waiting to unload, or a queue of people who have got off the buses waiting to leave the bus station). The Packet Delay Variation is a measure of how late a packet is liable to arrive, and defines the minimum buffering delay that will ensure reliable transfer of the audio samples.

To keep packetisation delay small, the packet size needs to be small, particularly if only one or two audio channels are being carried. Typically, when conveying an AES3 stream, a 48-byte packet would carry six audio samples per channel, resulting in a packetisation delay of six sample periods.

The time spent waiting to be output on each link can be kept small by reserving periodic 'slots' on the output, and by giving audio priority over other traffic. If the network allows large data packets to be conveyed in one piece, an audio packet that arrives at an output just after a large packet begins will have to wait until the end of the packet.

Packet routing within the switch should be a simple, fast, process that can begin as soon as the first few bytes of the packet have been received.

Buffering delay is minimised by a service that delivers packets at regular intervals, rather than allowing them to become bunched together.

We know what a live digital stream is because we use AES3 or MADI for digital audio data and both are designed to emulate simple analogue lines. An analogue line has no appreciable delay other than the speed of light. With all digital audio, there is an unavoidable delay for analogue to digital conversion, and another minor delay needed to reclock data, as it arrives, to the receiver clock. This delay is a one-time cost and normally doesn't change as the distance of the communication increases. The data on the cable is synchronous; like the analogue line, the data moves at the speed of light.

Assuming that the A-DC/D-AC delay is acceptable, professional streams such as AES3 and MADI behave very like an analogue connection and are happily used for live applications. It would be nice to have a similarly connection-oriented approach for professional applications, and it is possible.

Telecom companies have been using high capacity digital data trunks for years. These are designed to work with maximum efficiency and with minimum delay because their core application is telephone calls. Today's common-carrier structures are built on synchronous multiplex structures such as SDH and SONET. These are used in a near raw state to provide phone calls using data converted from analogue phones and with ISDN data more directly. These are also used to carry the packets of data generated by the Internet. An alternative network protocol could be useful.

Asynchronous Transfer Mode (ATM) is not a new technology. It was developed as a successor to early Ethernet LANs. It provides a more efficient use of the available network bandwidth but, more importantly for us, it provides guarantees of quality of service (QoS) and can handle all of its bandwidth more efficiently. The unprecedented success of the Internet and Internet Protocol eclipsed this technology for a time, but the technological arguments for its use remain particularly strong for real-time applications, such as video and audio streaming, where timing and bandwidth matter.

Like Ethernet, ATM sends data in a series of blocks, the difference lies in the size of the blocks and the way they are routed from source to destination.

ATM cells (not 'packets') are only 53 bytes in size. This makes them fast to pack and fast to unpack, reducing packetisation delay to a minimum. The address header is much smaller than an Ethernet packet because the route is negotiated at the beginning of a transmission, rather than independently for each packet, setting up a virtual circuit rather like a phone call but with whatever bandwidth you require up to the available limit on the ATM link. This keeps the transit time through the network to a minimum. Because the virtual circuit is continuously available, cells can be transmitted at a steady rate so the size of the receiving buffer can be minimised. The act of setting up a virtual circuit from a source to a destination is directly equivalent to routing a conventional audio signal; routing from a single source to multiple destinations is similarly straightforward.

ATM as a raw technology isn't quite enough. To handle professional audio adequately, and interchangeably, we need to define how an ATM stream will be used. This is what is defined in AES47.

AES47 provides for virtual audio circuits that have a guaranteed Quality of Service. Once the bandwidth of the digital audio signal has been set up, that bandwidth remains completely available until the circuit is disconnected.

A number of modes are possible, including a range of multichannel options covering 5.1 surround and multiway feeds. Perhaps the most useful mode for interchange purposes provides a transparent path for an AES3 signal, including all its associated ancillary data.

Using AES47, you will be able to make connections with communications delays that are very small compared with A-DC/D-AC conversion delays so you can use them for time-sensitive production applications such as music and two-way dialogues. The routing provides the flexibility you would expect but with the capacity to carry full-bandwidth uncompressed audio data. While the equipment at each end of the chain will need to be locked to an audio sample clock, the network really doesn't care. It is possible to have two audio streams at different sampling rates sharing the same physical cable, important for mixed media installations where 44.1kHz sampled audio needs to share routing with 48kHz signals.

AES47 coexists with other ATM data connections over the same network: video and talkback, for example. Perhaps ironically, IT data can also be carried in an ATM stream; in effect an Ethernet LAN can be piggy-backed onto ATM.

Imagine a single network connection capable of handling all the data links for a radio, TV, or real-time music production.

Details:

www.aes.org/standards