ATM for Video and Audio on Demand

David Greaves.
University of Cambridge and ATM Ltd.
email: djg@cl.cam.ac.uk
and
Mark Taunton.
Online Media.
email: mtaunton@omi.co.uk

Presented at the AES conference in London 25-Mar-96

ABSTRACT

For many applications, audio is now conveyed digitally. Video applications are following fast, particularly for video-on-demand. These digital streams require constant-rate digital channels between the source and destination, but a great number of all networks being planned today are based on the packet switched Asynchronous Transfer Mode (ATM) technology. This paper serves as an introduction to ATM, with emphasis on its ability to carry fixed rate channels, and describes the approach of the Cambridge Digital Interactive Television Trial, where Video and Audio on demand are transported to the Home over ATM.

1 ATM NETWORKING

An ATM network consists of an arbitrary mesh of links interconnected by ATM switches. The links may be of a mixture of speeds, generally from 2 Mbps to 2.4 Gbps, with a speed of 155 Mbps being one of the most common. The switches may be privately owned by one or more concerns or may be owned by a public telephone company which is offering a wide-area ATM networking service.

ATM networks were first seriously promoted by the ITU (or CCITT as it was then) in the late 1980's. ATM was seen as a single technology which as well as being a replacement for telephone switching equipment could also serve the growing world data traffic. The ATM Forum, started in about 1990, was a consortium of computer companies who saw ATM as a good technology for private networks. The Forum has successfully fostered an ever growing number of companies which offer ATM LAN technology for the office and corporate site interconnection. Today the excitement for ATM is focussed on the residential access and in-home areas.

An ATM network may be defined as one which carries ATM *cells* from one end station to another along a *virtual circuit*, in order, with a known *quality of service* or QoS. We will explain these terms, briefly, but for full details, the reader is referred to one of the books, such as Davidson [1].

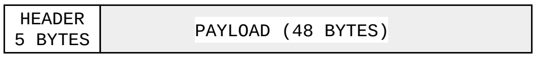

Figure 1: ATM Cell Format

Figure 1 shows that an ATM cell consists of a 5 byte header and a 48 byte payload. The 48 bytes of the payload are carried from the source end station to the destination end station unchanged by the network. The contents may be any type of user data and are not examined by the network. User data for a particular application must be arranged into a stream of 48 byte payloads by an ATM Adaptation Layer (AAL). The five byte cell header contains routing information which is handled by the switches, with various fields being changed at each switch, so that the limited number of circuit identifiers representable in the cell header, 2^{24}, may be spatially reused across the planet.

A virtual circuit is a fixed route which is established when an ATM call is created between one end station and another. All cells on the virtual circuit follow the same route. Also there may be multiple virtual circuits between two end stations, each concerned with a separate flow of information, but these separate virtual circuits do not necessarily follow the same route as each other.

For each virtual circuit there are two directions of cell flow, although the QoS for the two directions may be different. Very often, there is a primary or forward flow of cells for which the QoS is important, and a reverse flow which is only used for occasional control cells, whose bandwidth consumption may be typically neglected. An example would be a very simple video on demand system, where the forward path carries video and the reverse carries control messages to start and stop the stream.

ATM circuits can carry fixed rate data or variable rate data. However, all sources are expected to introduce data at rates below or equal to those specified when the connection was established. Cells transmitted at rates above this may be dropped by the network. A commercial network will have *policing* agents to detect offending cell streams and explicitly drop the cells, rather than allow them to interfere with the QoS of other users' circuits. For video on demand applications, a "push mode" of source rate pacing is normally used, where data is read from the server disk at the correct speed for real-time playback in the set top box (STB).

1.1 Quality of Service for ATM

The QoS of an ATM virtual circuit is normally specified using the parameters: the peak bandwidth, the mean bandwidth, the burst tolerance, the cell delay variation tolerance and the cell loss ratio. The peak bandwidth sets the minimum inter-cell spacing that a source is allowed to use while the mean bandwidth sets the average spacing (or mean rate of the cells). The burst tolerance essentially sets the period over which the mean rate is averaged, or it may instead be viewed as the maximum number of cells that can be sent as a burst at the peak rate, provided it has been sufficiently long since the last burst that the mean bandwidth parameter is not exceeded. The cell delay variation tolerance is an upper limit on the jitter that the network is allowed to introduce on the flow of cells. Jitter is defined as a variation in intercell spacing compared with the spacing used at the source. The cell loss ratio is the maximum fraction of cells that it is expected that the network will lose or drop.

The QoS of a virtual circuit is ensured by the ATM network using a number of mechanisms. One is bandwidth allocation and reservation, which operates at call admission time: when a virtual circuit is established or 'opened' between two end stations, the required QoS is given by the opening station. The network determines whether the QoS can be met by any possible route beteen the source and destination and if so chooses one, otherwise the call is blocked. The network also checks whether the requested connection can be provided without degrading the QoS of the virtual circuits already using the switch and link resources which will be shared with the new connection. A mechanism available to help here is priority within the switches. The switches contain routing tables with one entry per virtual circuit, which as well as the new header and output port number for that circuit, also hold the QoS requirements. Circuits which can tolerate greater delay variation, typically data, are given low priority within the switch, meaning that the switch will keep the cells in queue inside the switch when other higher priority traffic is contending for a switch output port. On the other hand, constant rate traffic, carrying music, video or telephony, is normally given the highest priority in the switch, since these services benefit from low delay through the system and will not tend to bunch up and queue owing to their fixed spacing at the source.

QoS Parameter	Value
Mean bandwidth	2.048 Mbps
Peak bandwidth	3 Mbps
Burst tolerance	2 cells
Cell delay variation	1 cell time
Loss ratio	1 in 10^7

Table 1: Typical QoS parameters for a 2.048 Mbps VOD stream.

1.2 Jitter and QoS for Video On Demand

For a 2 Mbps Video On Demand (VOD) stream, the QoS parameters would typically be as shown in table 1.[1] From the ratio of peak to mean bandwidth and the value of burst tolerance, the network could tell that the traffic is likely to be at nearly constant bit rate. The given values allow small variations in cell spacing (jitter) introduced after passage through a number of switches to be tolerated by a downstream policing agent. (Jitter producing a smaller gap between two cells will map into a higher measured peak bandwidth). The end station may actually transmit at any speed lower than defined by the QoS it has requested. If it is likely to do this regularly, then it should have requested a VBR (variable bit rate) connection at connection establishment, rather than a CBR connection (constant bit rate).

The jitter encountered by a CBR stream as it crosses the ATM network depends on the contention it encounters at the output port of each switch from other traffic. If it is the only stream, or the only stream at the highest delay priority, it will experience the minimum possible jitter which is one cell time peak-to-peak. This is because it must synchronise itself with the outgoing cell frame structure on the link.[2] If there are other streams of the same CBR priority at the switch port, queuing can occur. If the cells have essentially random inter-arrival times, as is the case if they come from uncorrelated sources, then the queuing delay depends on the number of streams, the percentage output port loading by this class of traffic and the output queue service discipline. For 90 percent loading with first-come, first-served service, the mean queuing delay will be about 10 cells, and the 99.9th percentile will be about 100 cell times. For reference, the intercell spacing for a 2.048 Mbps payload rate will be 74 cells and their mean rate will be 4.8 kHz. The tail of this jitter distribution is the important factor, since, as explained in section 3, the replay buffer must not under-run when the cell flow has one if its larger gaps. With round-robin service, where each VC contending for an output port is service in turn, CBR greatly benefits: each virtual circuit is bound to be serviced in inverse proportion to the number of active connections, giving a service time upper-

[1] Note that the loss ratio is typically a network wide parameter for constant rate traffic, rather than a per virtual circuit parameter.

[2] For unframed links, such as the Forum's ATM100 and ATM25.6 standards, the synchronisation delay is also the one cell time if the port was previously busy with another cell.

bounded by the the cell interarrival time.

When all of the links that the VC flow over are of the same speed and traffic mix, then the jitter may be expected to grow on a 'power basis', that is, in proportion to the square root of the number of switches. However, in the video-on-demand application, there will typically be a number of switches operating with 155 Mbps links and then one or two lower rate links in the access network near the home. Also in the VOD environment, the sources may be synchronised, either by means of the synchronous residual timestamp (SRTS), as explained in section 3.1 or because they have come from a common file server. In this case, there will be little random behaviour in the cell queueing and both the mean delay and the jitter will be very much reduced.

As said, the leaf links to the homes will usually be at a lower rate than the core network links. One technology for an ATM feed to the home is a 25 Mbps channel encoded into a conventional 8 MHz cable TV slot by ATM CATV modem technology [2]. Another technique is Carrierless Amplitude and Phase modulation (CAP-16) [3] bringing 51 Mbps to the home from a kerbside ATM cabinet over the exisiting copper telephone pair. In the Cambridge iTV Trial, and other places, we are using baseband modulation at 2 and 25.6 Mbps of the point-to-point coax from the home to the kerbside unit. Directional couplers (hybrids) are used to give a duplex ATM link.

The lower rates of the leaf links means they have longer cell times, and therefore jitter in terms of microseconds for the same type of queueing is increased. These links also tend to have lower numbers of contending streams, and so jitter will be reduced. Finally, owing to the high cost of 'last mile' bandwidth, it is desirable to run these links at very high utilisation, perhaps leaving only one percent non-CBR traffic for control cells of various types and to accommodate link speed tolerances. This level of utilisation requires per-virtual circuit round-robin queuing on the leaf switch output port. Such queuing limits jitter to one intercell arrival time.

A back-of-the envelope calculation of the jitter over a VOD ATM network when highly loaded with CBR traffic gave the following result: If there are 5 switches between the server and the leaf switch, using 155 Mbps links, first-come first-served queuing and 90 percent loading, then a 2 Mbps channel will experience about 500 μs 99.9th percentile jitter. The jitter in a leaf switch using round-robin queuing for each VC, if it is fully loaded, will be the same as the intercell spacing, which is 200 μs. Therefore a replay buffer with nominal fill of about 10 cells (or two MPEG TS AAL-5 frames) is appropriate. This introduces much less than a video frame time of delay and requires insignificant RAM memory, when implemented in the host DRAM of the STB.

Preliminary measurements of the jitter at the STB on the trial confirm a 700 μs peak-to-peak jitter. We hope to report further measurements of the Trial Network at the conference.

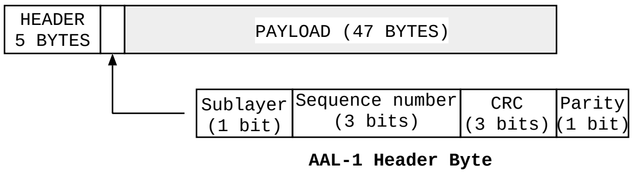

Figure 2: ATM Class 1 Adaptation Layer

2 ADAPTATION LAYERS

The ATM network can in principle carry cells with arbitrary payload structure, but in practice two standard ATM adaptation layers are in common use [4]. The first is AAL-1, which offers a fixed rate byte stream over the ATM network, where the output rate of bytes exactly matches the input rate. The second is AAL-5, which is used for data which is sent in blocks known as *datagrams* typical of data communications technologies such as X.25, Ethernet and the IP protocol of the Internet. AAL-5 may also be used for CBR traffic (and is in the Cambridge Trial) provided the datagrams are generated periodically. Of the other AALs, AAL-2 is not fully defined, but is envisaged for variable rate media streams, as produced by VBR video codecs. AAL-2 may become important in the future. AAL-3/4, which is a cojoining of the older AAL-3 and AAL-4 is also for computer data, but was rejected by the computer industry when AAL-5 was defined.

The format of data using AAL-1 is shown in figure 2. Here, the first byte of each cell is used by the AAL for protocol information. Half of the 8 bits are used as a CRC and parity check of the other half, resulting in only 16 valid byte values. The data portion is used as a 3 bit sequence number, which is incremented for each cell transmitted, allowing the destination to detect missing cells, and there is one remaining bit, that conveys the 4 bit SRTS value in bit sequential form over successive cells. SRTS is explained in section 3.1.

The format of data using AAL-5 is shown in figure 3. Here the variable length datagram which is required to be transmitted is segmented into the appropriate number of cells. The last cell will in the general case not be full, and contains a pad field and a trailer. The pad field is just to round up the overall length to an integral number of cells. The trailer contains a length field (of the valid data) and a 32-bit CRC of the whole message. The last cell is indicated using a bit in the ATM cell header known as the ATM User-to-User information element, which is not changed by switches from one end to the other, and really acts as an extension to the payload.

3 CONSTANT BIT RATE TRAFFIC AND TIMING RECOVERY

For real-time streams, such as video and audio on demand, the data must be delivered to the replay codec at

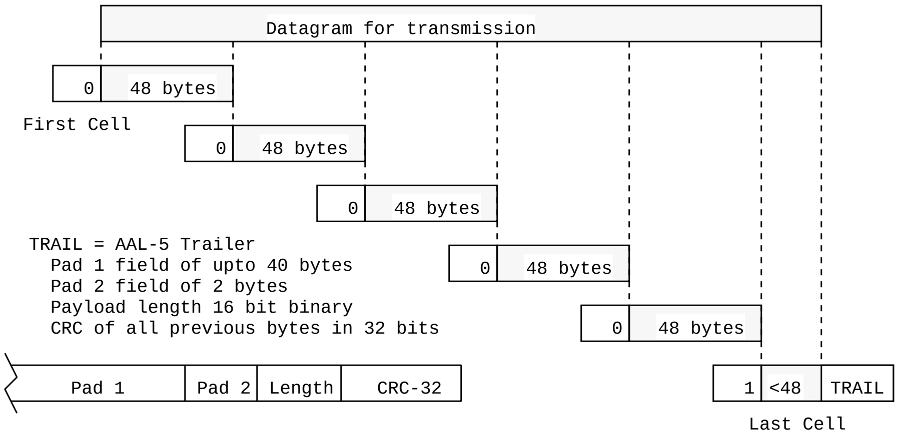

Figure 3: ATM Class 5 Adaptation Layer

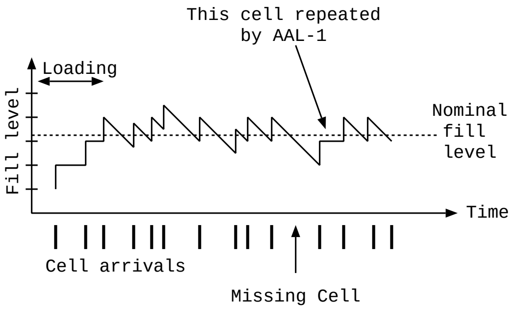

Figure 4: Graph of CBR dejitter buffer fill level

the correct speed, which is exactly the speed the media is being captured or replayed at its source. Using simple crystal oscillators at the source and destination is not always sufficient, since with typical accuracies of tens of parts per million (ppm), one byte in every 10,000 or so will have to be dropped or repeated, causing perceptible perturbations to most applications, especially those using compression. However, with suitable adjustments (dropping or repeating larger groups of bytes, in a controlled manner), this simple mechanism can be successfully applied; it is currently used in the Cambridge Trial (details in section 5.4).

For CBR over an ATM network, source clock recovery is usually used, but this is more complicated over ATM than a circuit-switched channel, since a cell may be dropped by the network or it may be late to arrive. The solution is to use at the receiver a playout/dejitter buffer.

Figure 4 shows how the level of data will typically change within the playout buffer. The system starts in a loading phase, and accumulates data until a nominal level is achieved, then it commences play out. During playout, the input and output average rates should be equal (when using the mechanisms explained in section 3.1) and so the fill level will only vary from the nominal level as a result of jitter in the arriving cell stream. If the nominal level is too low for the level of jitter, then underrun occur, and the system will reenter the loading phase. If it is too high, undue delay is introduced. Delay is undesirable for interactive services.

Also shown in the figure is the action of AAL-1 under single cell drop. Upon arrival of the cell after the dropped cell, the system either repeats the last cell or else introduces dummy data to restore the correct playout level. The errored data will not be rendered by the more sophisticated decoders for audio and video, since these accept a signal indicating that the current data is a substitute for missing data and so invoke their error concealment mechanisms (normally interpolation of some sort). AAL-1 can detect and handle up to 8 cells being dropped, a most unlikely occurrence for CBR traffic.

Not shown in the figure is the handling for severe error cases: if the level ever drops below a lower limit, or becomes zero, then the system will reenter the loading state, and if the level goes above an upper limit, then cells are discarded until the nominal level is obtained.

3.1 Clock Recovery from CBR Cells

There are two main techniques for regenerating the original constant bit rate clock at the receiving end station of a CBR flow: adaptive clock and synchronous residual timestamp (SRTS).

As shown in figure 5, adaptive clock uses a depth threshold on the replay dejitter buffer to control the rate of a local voltage controlled- or numerically-controlled oscillator. Data is read from the dejitter buffer at the rate generated by the oscillator. The cell gate implements the initial loading algorithm, the interpretation of AAL-1 headers and the insertion of dummy data under cell loss. Adaptive clock can also be used with other AALs, and in particular is often applied for AAL-5 datagrams which contain MPEG transport streams as described later.

SRTS requires a spanning set of the physical links of network to carry an 8 kHz reference clock from a master station to all participating end stations, as shown by the thicker lines in figure 6.

The source clock must have a nominal rate which is

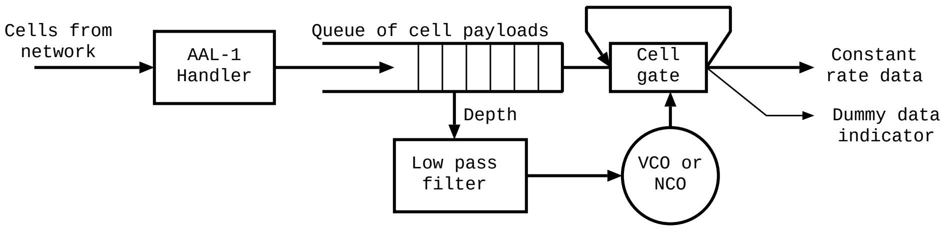

Figure 5: Timing recovery using the adaptive clock technique

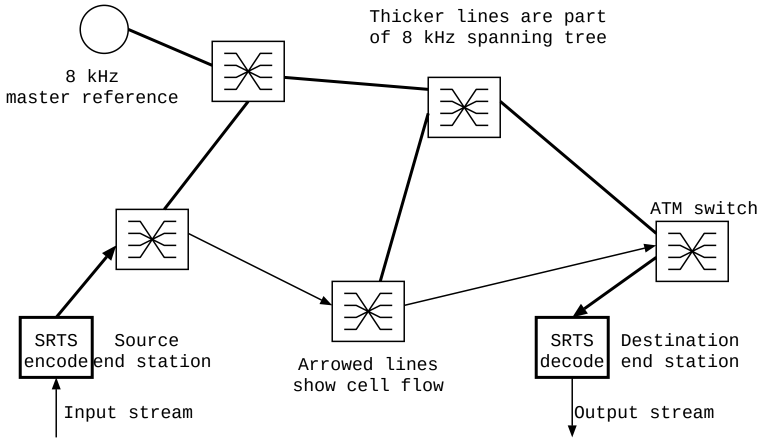

Figure 6: Cell flow and 8 kHz clock distribution for SRTS timing technique

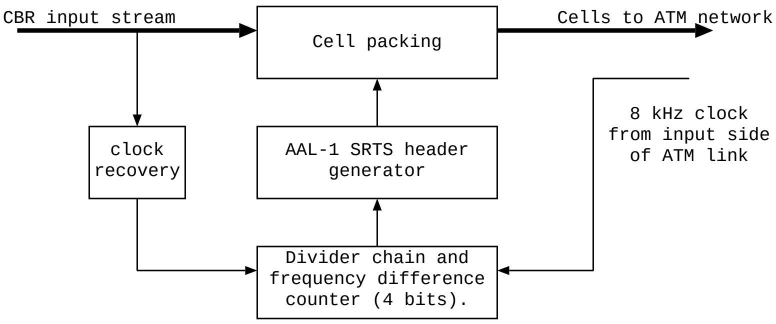

Figure 7: SRTS encoding process

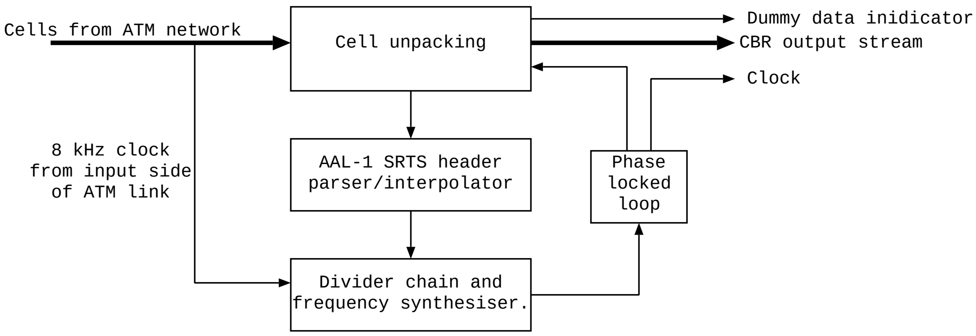

Figure 8: SRTS decoding process

known by both end stations, but it will also have a small, perhaps slowly varying frequency offset. This offset is transmitted to the other end using the SRTS values in the AAL-1 header, so that the receiving end-station may exactly resynthesise the source clock rate.

The details of the encode and decode process are show in figures 7 and 8. At the source end station, the CBR source clock is extracted from the data and compared in a counter and divider arrangement against the network 8 kHz clock. The appropriate division ratio has been hard-wired or preset using the prior knowledge of the nominal source clock frequency. The error is conveyed in the AAL-1 header to the receiving end station using the four bit SRTS value. This is capable of handling +/- 200 ppm of frequency offset. At the decoder, the headers are extracted from the cells and used to resynthesise the input clock rate from the 8kHz reference, which of course is also available at this station. A phase-locked loop of a few Hertz bandwidth is generally used to reduce the jitter components in the resynthesised clock.

In general the SRTS offset will not be exactly one of the values represented by the set of 16 directly conveyable, and the SRTS value will oscillate between two adjacent values with the appropriate duty cycle. These oscillation cycles may last many seconds and so lie within the passband of the PLL. This causes a small amount of very low frequency wow [5].

A possibility in the future is to use video servers which synchronise themselves with the network 8 kHz clock, in which case the SRTS value inserted in the cell header can always be zero.

4 MPEG TRANSPORT

MPEG (the Motion Picture Experts Group) is a formal Working Group of the International Standards Organisation. As part of its work on compressed audio/visual content representations, MPEG has defined a standard (ISO/IEC 13818-1) for the carriage of a multiplex of audio, video and other digital data, over generalised communication paths. The standard defines two types of multiplex: Program Streams (designed for use in situations where reliability is high), and Transport Streams (TS), which are intended for environments where high reliability is not guaranteed. Transport streams are now widely used for carrying audio/visual and other stream-oriented data.

An important aspect of the MPEG transport stream standard is that it allows for many individual mono-media streams to be carried in a multiplex. These streams can be independent, or may be logically grouped into related sets, forming complete programmes. Such groupings are identified by special programme association tables and programme mapping tables, which are also carried in the multiplex. Within a group there may be multiple instances of a particular type of stream. Most commonly, a complete audio/visual programme might be carried as a single video stream, but having multiple audio streams associated with it. An example of this structure in action would be on board an international airline flight, where invidual passengers might be watching the same movie, but could each hear the soundtrack in their own native language.

In a transport stream, the data is divided up into a sequence of packets of fixed length. Data for different streams, even if parts of the same overall programme group, are always contained in separate packets: this greatly simplifies the processes of multiplexing and demultiplexing. Although the underlying protocol layers are deliberately left unspecified, the standard has been consciously specified in a way which fits well with ATM networking standards. Transport streams may be carried over an ATM network using either AAL-1 or AAL-5. The MPEG TS packet length is 188 bytes, and so two MPEG TS packets fit exactly into the available payload space of an AAL-5 datagram of four ATM cells. Thus the complexity of mapping between protocol layers is reduced to a regular subdivision of data into successive fixed size units before onward transmission, and data space wastage is eliminated, because no padding bytes are needed to fill the last cell.

Time is a very important element in the transport stream concept: each TS packet can optionally contain timing information, relating to the particular stream, part of whose data is contained in that packet. This is essential to allow synchronisation, both between associated stream data (audio to video - i.e. lipsync) and between the decoder and the transmitter or encoder. Without this - since in general the timebases of the two ends are independent - there would be serious difficulties for reliable presentation of the content to the ultimate viewer.

A transport stream multiplex is intended primarily for unidirectional operation. Reliable transmission of the data end-to-end is neither guaranteed nor assumed (since the introduction of mechanism to guarantee reliability would potentially have a huge impact on both latency and intermediate buffering costs), but typical link error rates are low enough that this is not considered a problem for most real-world use of transport streams. Additionally, the individual MPEG video and audio streams, as well as the transport stream format, include various mechanisms such as CRC (cyclic redundancy checking) to allow error detection and some degree of error recovery and/or concealment.

5 CAMBRIDGE DIGITAL iTV TRIAL

The Cambridge Digital Interactive Television Trial started in September 1994, with ten users connected in the first phase. ATM networking technology is employed throughout the system, with two-way digital data being carried alongside cable TV signals over fibre optic links. In phase two, begun in March 1995, the network and the content server facilities were substantially enhanced, and

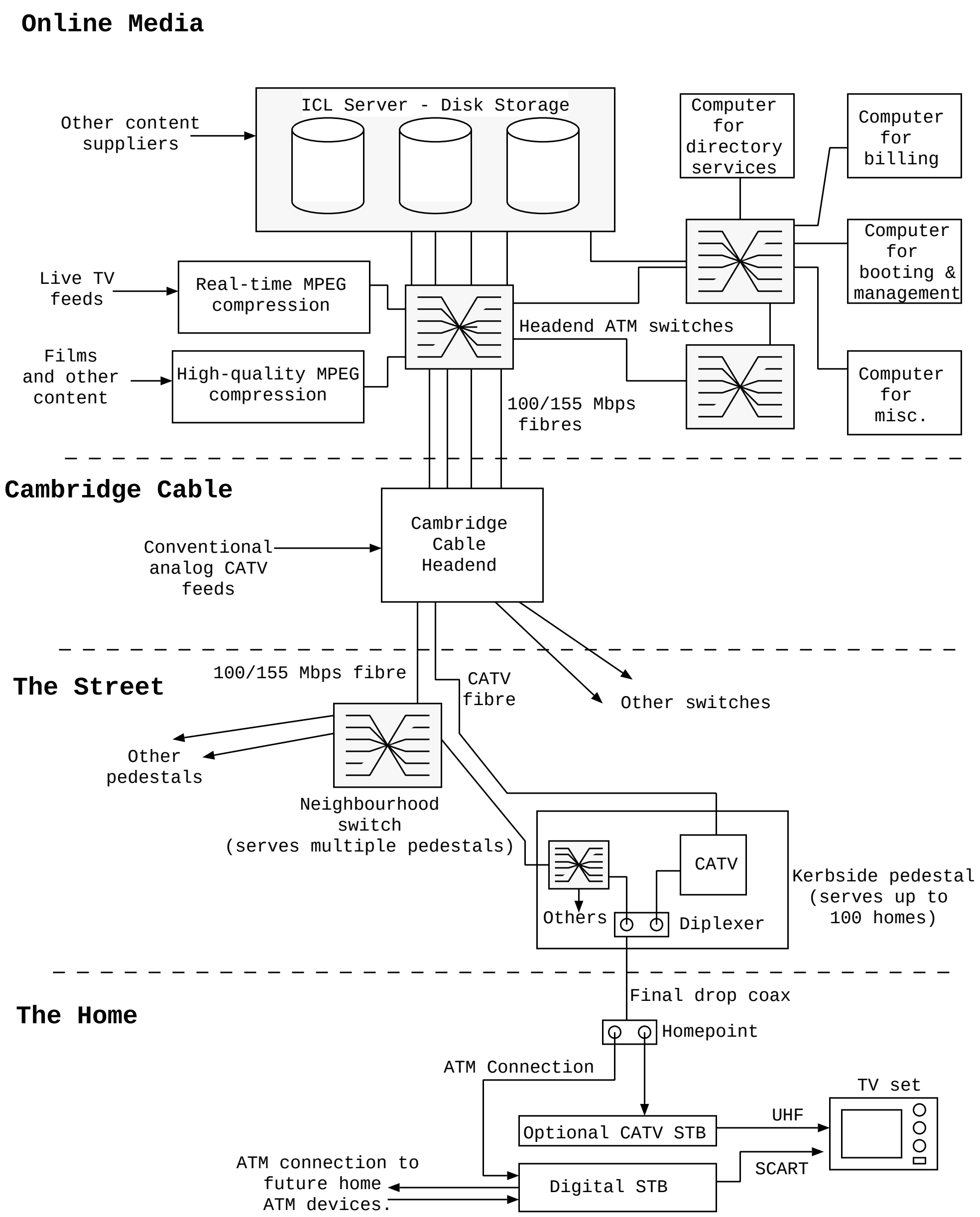

Figure 9: Cambridge iTV Trial network configuration

the number of user sites (homes, schools and businesses) was increased: by January 1996 it approached one hundred.

The partner companies in the consortium which provides and operates the trial infrastructure - the various hardware and software components and the network links - are Online Media, Advanced Telecommunications Modules Ltd (ATML), Cambridge Cable, ICL, and SJ Research. The trial headquarters are located in the offices of Online Media in Cambridge. A second group of companies is responsible for delivering the various services and content accessible via the system. These include the BBC, National Westminster Bank, ITN, Tesco, IPC Magazines, Acorn Computers, and Anglia Television; other organisations involved in the service consortium are BMP DDB Needham, the Post Office, NOP (National Opinion Polls), and the Independent Television Commission.

Facilities presently available to participants in the trial range from home banking via the TV, through software programs for both education and entertainment, general information services, and of course interactive video and audio access on-demand. The last includes both dynamically updated content - news and weather reports, and other topical material such as programme guides, with TV trailers viewable on-demand - and an expanding archive of general interest programmes such as comedies, drama, documentaries and music concerts. Contributors to the audio/visual services include the BBC, ITN and Anglia Television.

5.1 Trial Architecture

The system architecture, shown in figure 9, starts with a high capacity ICL server at the "digital head-end", with content stored in a array of hard discs attached to a modular parallel-processor complex. This can deliver a large number of independent digital data streams, through the ATM switching network, to individual digital set-top boxes (STBs) where the data is decoded and presented via TV and/or audio equipment. The maximum number of simultaneous streams is determined by a number of factors, including the data rate used for each stream, and the configuration of the server. If fully populated with discs and processors, the ICL server would be able to provide as many as 3000 continuous video streams. The current system, which uses fewer than the maximum possible processing elements, is well able to support all of the client STBs so far installed. Of course not all possible users are ever active at the same moment, and so, as with the public telephone system, a larger number of STBs can be connected than could all use it at once. Thus there is considerable room for further growth. A single server of this type could in the future support ten thousand or more STBs.

The ATM switching system serves to carry both the dynamic requests of active users, and the associated responses from the server (individual messages, or real-time programme streams). In the current infrastructure, each physical link between an STB and the network can sustain a maximum data rate of 2 Mbps in each direction simultaneously. For the present, almost all data traffic follows the "client/server" model, with messages being carried from an STB to the server, and replies returned over the reverse route. Of course, the most obvious example of such data is the audio/visual digital programme content fed out by the server for the user to watch or listen to. However it is quite feasible through the ATM network for STBs to arrange to communicate directly one to another, independently of the central server. Thus, fully two-way symmetrical services (e.g. video-phone) are in principle possible, although this would need additional hardware - such as a video camera and a digitiser/compressor - to be attached to the STB.

The hardware of the network comprises firstly a number of switches connected directly to the server's multiple data ports; these switches are supplied by ATML. They direct the ATM data traffic over fibre optic cable between the server and Cambridge Cable's network distribution centre, which is located 1.5 miles away. From there the network branches out over more underground fibre optic links, towards the various user sites.

The final staging post on the journey for each piece of data in transit, is one of the familiar kerb-side cabinets, which contain the equipment used to supply cable TV signals to subscribers' homes. Alongside the standard CATV distribution box in each cabinet used by the trial is added an ATM switch built by SJ Research. This serves two purpose: firstly, as a switch, it routes messages to and fro between the single network (fibre-optic) link to the head-end, and the multiple individual (copper) links to nearby user premises. Secondly, it electrically combines the bi-directional digital data streams, used for interactive TV, with the conventional CATV feed, onto the same standard coaxial cable already present to carry the cable TV signal to the home. The diplexing circuitry is simple, consisting of no more than filters, amplifiers and isolation components: no complex modulation scheme is employed and the digital data are carried effectively as a band-limited baseband signal extending in frequency up a few MHz. At the other end of the coax "final drop" is a wall box, usually mounted inside the house or school building, which performs the reverse function to the kerbside diplexer. It separates the signals on the coax into the CATV feed, made available at a socket for connection to the cable TV control box, and the bidirectional digital data, which is linked by UTP (Unshielded Twisted Pair) wiring to the digital set-top box.

5.2 Set Top Boxes

The STBs on the trial are provided by Online Media. As shown in figure 10 they contain a number of elements, including a central processor, memory, the network interface circuitry which is fitted as a plug-in module, and the audio and video decoding hardware. The STB is remotely controlled, by means of a separate Infra-Red (IR)

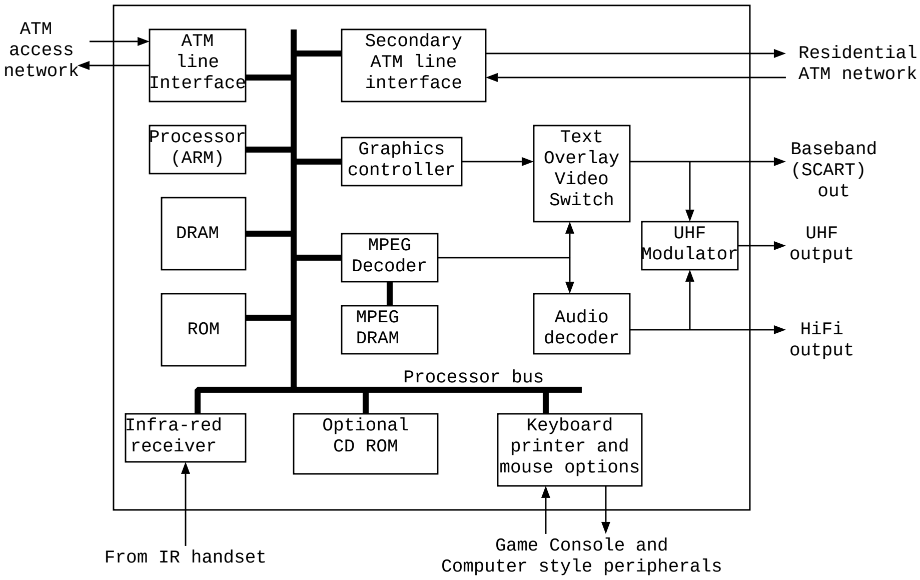

Figure 10: Components within a Set Top Box

handset.

Currently MPEG decoding of both video and audio is performed using commercial dedicated hardware devices; however, the potential for less costly software-based decoding is being actively pursued. The ARM central processor used in the current generation of STB is capable of performing real-time MPEG audio decompression in software. However, for reasons concerned with the arithmetic precision available, the final signal quality is slightly lower (i.e. has a lower signal-to-noise ratio) than that produced by the dedicated chip. Additionally, the workload involved in performing the various data flow and system management functions in the STB is such that the amount of processing power left over for audio decoding may not always be enough. Future designs will increase the processing power substantially and are intended to permit both audio and video decoding to be performed in software alone.

The decoded audio and video outputs are made available at the rear of the STB in a number of forms. SCART connectors provide baseband video and stereo audio in the now-common arrangement; additional sockets present both composite and separate (Y/C format) baseband video signals and a stereo pair for audio, and finally, a modulated UHF signal is output, which is suitable for direct connection to the TV aerial socket. Daisy-chaining of both SCART and UHF signal paths allows the STB to form as part of a more complex A/V system: the STB can disengage itself (e.g. on entering standby mode) and allow signals from a VCR, or an aerial or cable TV feed, to pass straight through to the TV. This design appears somewhat complex, but for the purpose of the trial was viewed as necessary, in order to give maximum flexibility to suit the varying needs of the trial participants.

Around 40 percent of the users on the trial also have stereo audio equipment of some sort near their STB/TV installation. Of these, over half have made or intend to make a connection for the audio signals, so as to obtain best quality sound. Some of the users are active audiophiles; one has gone so far as to connect an oscilloscope to the STB in order to view the audio output signal!

5.3 Encoding standards

The audio and video data used in the trial are all generated and transmitted in accordance with international MPEG standards. Currently, the individual audio and video streams are encoded using the respective MPEG 1 standards. Most video material is digitised at a resolution of 320 pixels horizontally, by 288 lines vertically, at a 25Hz frame rate (MPEG 1 compression manipulates only frames, not interlaced fields) for display on PAL TV. As in broadcast TV standards, the chroma (colour) information is sub-sampled and carried in a lower resolution form; this works because of the eye's lower sensitivity to rapid changes in colour over local regions of a displayed image. The total effective data compression ratio (relative to a raw digital representation of a video sequence at the same luminance and chroma resolutions) is in the range 20:1 to 30:1.

The audio content is normally sampled at 44.1kHz (and always in stereo), although 32kHz and 48kHz sample rates are also possible. It is then compressed using hardware to the MPEG 1, layer II format. This is a MUSICAM compression scheme, using psycho-acoustic modelling to enable the discarding of parts of the original signal which would not be perceived in any case, and also allowing coarser quantisation of the signals in parts of the audio frequency range where the associated noise is less discernible. Compression ratios supported in MPEG 1 (relative to 16-bit linear digital source material) lie in the range 4:1 to 20:1, with a typical ratio of perhaps 6:1 or

7:1.

The encoded data must ultimately be transmitted over a network which has a finite maximum data rate - currently 2 Mbps. This imposes constraints on the compression process, to ensure that the final bit-rate is within this limit. In fact there are certain additional factors in the overall data transport system (such as the ATM cell headers, and other protocol costs) with the result that the net bit-rate available to carry the actual compressed data streams is around 1.6-1.7 Mbps. The MPEG 1 audio and video compression standards were aimed at encoding for a total compressed data rate of around 1.5 Mbps, and this is therefore a good match to the current trial infrastructure. The encoders used allow the video and audio compression ratios to be separately adjusted, to suit the total bit-rate requirements. After encoding, the audio and video data is combined into the MPEG TS multiplex and held in this form on the server, ready for transmission.

For replay, since we are using CBR, the server need only examine the bit-rate information provided in the stream once, at the start, and thereafter sends ATM cells at a fixed rate, calculated to deliver (on average) the requested bit-rate over the network. This is the source rate pacing.

5.4 Stream synchronisation

A feature of push-mode source rate pacing is that synchronisation between the server and the decoder is a one-way process. This is clearly useful for multicast and broadcast streams. Given that the source rate clock has been recovered from the ATM stream (using either SRTS or adaptive clock), there remain two main approaches that the decoder can use to match its overall content presentation rates (output sample rate for audio, frame rate for video) to the CBR recovered clock, which reflects the average rate at which the corresponding compressed content is arriving.

In the first method, the STB arranges to synchronise its master clock(s) to the incoming bit stream (whose rate is derived from the transmitting system's clock), by extracting the timing information in those TS packets which contain it, and comparing the arrival time of such packets (as measured by the STB clock) with the timestamps. It can then gradually adjust its own clock (e.g. by slightly "pulling" a master crystal oscillator's frequency up or down via a control signal), so as to bring the local and received clocking values into line. In this context, the level and statistical distribution of jitter in the received data stream is very significant in determinig the time it takes to achieve (to within a specified margin) the desired timing lock.

The alternative approach is slightly simpler; individual blocks of content data - audio frames (1152 samples in MPEG 1 layer II audio) and/or whole video frames - are either presented twice, or discarded, whenever the compressed input stream data buffer level goes respectively below or above preset limits. This obviously gives rise to effects in the content presentation which are potentially discernible.

In the STBs currently deployed on the Cambridge iTV Trial, the second method is used. When the associated content discontinuities do occur, users may detect a slight stutter or jump in the sound, or a brief irregularity in otherwise smooth motion on video. However, most times when this happens the effect is usually found to be unobjectionable, and in any case such events are rare, given that the basic timing accuracy at each end of the transmission chain is reasonably high. The unit of content data which is dropped or repeated is much larger than one bit or one byte, so for a given timing error, the interval between occurences of disturbance scales up proportionately. Thus with a constant relative timing error of (say) 20ppm between source and receiver, and an audio frame time of 26ms, a single event (frame repeat or drop) will occur once in 20 minutes. Depending on the amount of silent or quiet material in the stream, only one in 3 or 4 such occurences may actually be noticed, leading to a perceived disturbance rate of around once per hour.

5.5 Video quality

If performed well, MPEG 1 video compression is capable of representing most video material to a quality level which is commonly found acceptable. The best examples of MPEG 1 compression have in general been produced from film stock. Because of the natural properties of film material as a source - in particular because it is represented as whole frames (so-called "progressive coding") - it is a good match to the frame-based MPEG 1 compression scheme, since there are no interlace-related artifacts.

The bulk of the video material available on the trial is sourced originally from TV cameras, and hence contains interlaced image detail. Certain limitations of the encoding scheme become more noticeable (though not necessarily offensive) in this case. Particular types of scene, such as more artificial environments with sharp edges and lines, can sometimes be seen to exhibit jagged patterns, especially on near-horizontal edges. Another issue is motion: certain types of movement are not well represented when the detail variation between two successive interlaced fields is lost or blurred during the frame capture process. Finally, the real-time encoders currently used to digitise and compress the data do not perform any sophisticated video pre-filtering to reduce the effects of noise in the original analogue signal. In some cases this gives rise to poorer results than a more elaborate and expensive decoder could achieve. However, they can perform well when given "clean" source material. For example, the TV programme trailers are encoded off high quality tape, supplied directly from the broadcaster, and some are perceived as visually better than analogue playback from a good consumer-grade VCR.

As yet, specific detailed user opinions on the video quality have not been solicited, but some subjective indications have already been received. Those trial users who have commented consider most digital video material to be visually somewhat poorer than broadcast TV. However, most users do seem accepting (or at least tolerant!) of the picture quality obtained. The network has been broadly quite reliable in use, and in consequence, problems relating purely to the digital transport of the data (such as lost or erroneous data causing picture or sound break-up) are rare. Most problems observed relate to the digital encoding of the data, as already discussed.

5.6 Audio Quality

Obviously, the decoding and analogue audio circuitry in the STB will have properties which may be reflected in user perception of the quality of sound output. However, a likely more significant general issue is the effect on the content of the MPEG audio compression system. There is a general inverse correlation between the degree of compression applied to the original data, and the achievable subjective audio quality level.

The lowest audio bit-rate which is used for audio content in the trial is 160 kbps; this represents a compression ratio of 8.8:1. Originally, this rate was used generally, for the sound on locally encoded video material, such as news and weather reports which are captured and encoded in real-time as they are broadcast, and also for material such as documentaries and nature programmes which are compressed "off-line". However, feedback from users led to the conclusion that this rate was insufficient, as certain types of noise and distortion in the sound (characterised as "graininess" or "grittiness") were sometimes observed, particularly with music material. Accordingly, a rate of 192 kbps is now used by default. This is a good compromise figure, since of course every bit used for audio is a bit which cannot be used for video, and vice versa.

In the special case of audio-only material (e.g. the many archived BBC radio programmes which are being encoded as an on-going activity), the bit-rate limit is not a problem, since all 1.7Mbps or so is available. In principle, the content could be sent as uncompressed linear audio (needing 1.5Mbps for 44.1kHz stereo 16-bit samples). However, for a number of reasons - principally the server storage space needed for an archive now running to many hundreds of hours of content - this course has not been followed. Instead, compression is still used, but at higher bit rates. The exact choice of rate depends on the material to be encoded: the highest normal rate (e.g. for music material) is 256kbps. For simple, essentially mono, speech material (e.g. book readings), rates may be cut back to as low as 160kbps - the quality problems noted earlier are not apparent in this case.

Subjective user views on the delivered sound quality are less well known than for video. However the previously-mentioned keen audiophile considers it to be generally "not bad", although following observation with the oscilloscope he reckoned it to be "a little dirty top and bottom" (meaning uncertain) and the tone was described as sometimes "rather thin". Further research is likely, in order to gather a more objective picture of user reactions as a whole.

5.7 Further developments

For the future, there are plans to experiment with increased data rates: the bottleneck in the present architecture is the 2 Mbps interface hardware at each end of the final link between the kerb-side cabinet and the STB. A number of home links are being upgraded to operate at the 25.6 Mbps rate, although perhaps this is a luxury that will not be available in many parts of the world, owing to the inferior quality cabling that will must be reused.

Once bit rates significantly over 2Mbps are available, it will be of interest to experiment with A/V data encoded using the more advanced MPEG 2 standards. In particular, the MPEG 2 video Main Level, Main Profile (MP@ML) definitions allow for video at a resolution up to 720x576 at 25Hz, optionally interlaced as two 288-line fields at 50 Hz, with bit rates up to a maximum of 15 Mbps. For various practical reasons - total network capacity, and the implications on storage space at the server - more typical video bit rates are expected to be in the range 3-6 Mbps. MPEG 2 also extends the MPEG 1 audio compression schemes, and in particular supports a form of multi-channel audio with up to 5 full-frequency channels and one low-frequency channel, so allowing for surround sound.

The STBs currently used in the trial are fitted with video and audio decoders for MPEG 1 only. However, a version of the STB capable of MPEG 2 decoding is currently in development by Online Media. A greater problem is that, particularly for video, new - and currently expensive - encoding hardware would be required in order to support the more complex MPEG 2 standards, and so this is not likely to be deployed generally on the trial in the near future.

As mentioned, the STB has provision for connection of external analog Hi-Fi. Shown in figure 10 is optional provision for extension of the ATM network into the home. In the future, this can be used to connect to digital Hi-Fi or extension LCD video display panels, or for the introduction of a video baby monitor where a stream of ATM cells from an ATM camera in the bedroom would be received by the STB for display at the touch of a button on the IR handset.

6 FINALLY

Although ATM was first widely promoted for voice applications and the cell size of 48 bytes was chosen as a compromise of efficiency (ratio of header to payload) over packetisation delay (48 bytes is 6ms of 8 ksps speech), it is not until recently that ATM has started to be deployed for telephony applications, and then only in advanced office environments where PC and phone integration is occurring. However, the trend to advocate ATM for all forms of multimedia transmission will continue, and digital audio over ATM is likely to become common.

Digital audio streams which carry a pair of stereo channels, such as SP-DIF can be carried over ATM AAL-1 without consequence, provided both channels are routed to the same destination. Circuits akin to that for the adaptive clock techniques are already widely used at the receiving end of an SP-DIF link. However, when separate speakers are fed by different ATM links, the sound must be split at an ATM switch, or at least multicast to the various speakers, each of which can render its own channel.

Now is the correct time to define standards for consumer and professional audio over ATM. There is a wide variety of options to solve the synchronisation problem, although none will change the fundamental way ATM operates, as outlined in this paper. There are also many other problems to be addressed, including piracy, low-cost charging for pay-per-listen, and standardisation of control interfaces.

About the Authors

Dr David Greaves completed his PhD at the University of Cambridge in 1989 on the subject of ATM metropolitan area networks. He was one of the first advocates of ATM technology for the private LAN market. Currently he is a University Lecturer in Cambridge and also Chief Scientist at ATM Ltd, a participating company in the Cambridge iTV Trial.

Mark Taunton is a Principal Design Engineer at Online Media (part of the Acorn Computer Group). After graduating B.Sc. (Hons) from the University of Edinburgh, he joined Acorn Computers Ltd in 1982. Since then he has been involved in a range of technology developments, from processor design through operating systems and compilers. Digital audio is one of Mark's long-standing special interests. He has worked for Online Media since its formation in June 1994, focusing particularly on software-oriented audio and video processing.

References

[1] "Broadband networking ABCs for managers: ATM, BISDN" Robert P. Davidson, New York: John Wiley & Sons, c1994.

[2] "Towards Management Systems for Emerging Hybrid Fibre-Coax Access Networks" S Ramanathan and Riccardo Gusella. IEEE Network Magazine, September 1995.

[3] "51.84 Mbps 16-CAP ATM LAN Standard" G Im, D Harman, G Huang, A Mandzik, C Nguyen, JJ Werner. IEEE J-SAC Vol 13, No 4. May 1995.

[4] "B-ISDN ATM Adaptation Layer Specification" ITU Recommendation I.362.

[5] "Timing issues of constant bit rate services over ATM" M Mulvey, IY Kim and ABD Reid. BT Technology Journal Vol 13 No 3 July 1995.