MPEG Systems Layer

MPEG video elementary stream not much use on own – maybe for playing silent video clips.

ES part of MPEG Systems Layer, separate part of MPEG standard

- provides means of combining/multiplexing video, audio, and ancillary data bistreams
- transmit program-specific info (PSIP) enabling navigation and access control
- provide framework for error protection
- convey timing information to decoder

Transport Stream

MPEG video, audio, and data streams usually combined together to form a Transport Stream. A Transport Stream consists of short, fixed-length (**188 byte**) packets concatenated together.

Each packet contains a header followed by a payload which is a chunk of a particular elementary stream. The header distinguishes what the payload/data is assigned to. The header consists of a (non-unique) one-byte start code which can be used for flywheel based synchronization, followed by a packet identifier (**PID**) which is a 13-bit code indicating which elementary stream the payload belongs to. What else have we done electronically that's would be labeled as flywheel based? A crystal oscillator, which needs frequent injections of energy at the proper time to keep running at the resonant frequency. Not like the start bit in a serial stream, since the serial stream isn't constant. This stream is constantly going.

Why have to have a PID? Not to distinguish audio from video, since audio and video are all part of a given unit. To distinguish to which subchannel the elementary stream belongs (e.g. 7.1, 7.2) since a TV station may be running multiple subchannels simultaneously.

The task of a transport stream multiplexer is then to form the elementary streams into packets, to buffer them and to select them for transmission according to the rates at which the elementary streams are generated.

For reasons partly connected with timing information, video elementary streams undergo intermediate stage of packetization into packetized elementary streams (**PES**s) at which stage some timing data is added.

Chart of Transport Stream and Structure

We saw this last time. Sync byte starts it, then some other things, then the packet ID (PID), then more header stuff, then the payload. What's in the payload? The elementary stream; this includes the headers associated with pictures, macroblocks, etc. From the Ennes workshop in Sacramento, the packets have a 4-byte header and 184-byte payload.

PSIP (Program Specific Information Protocol)

A Transport Stream may contain several video, audio and data channels (only one in HD since HD needs the room).

The decoder needs to know which video and audio signals go together to make each particular program. The MPEG Systems Layer provides a two-layer system of navigation through the Transport Stream.

First, packets with a PID of zero contain what is known as the **Program Association Table** (PAT). This is a list of programs together with the PIDs of further tables, one for each program. Each of these further tables, known as **Program Map Tables** (PMTs) contain a list of all the elementary streams making up the program and the PIDs identifying them.

The MPEG specification for program-specific information (**PSI**) also includes a **Network Information Table** (NIT) which contains information about the physical networks different programs are transmitted on (frequencies, for example) and a **Conditional Access Table** (CAT) which provides a framework for scrambling, encryption and access control, though the actual implementation of conditional access is not part of the MPEG specification. The stations send these tables along in the data stream.

If channel 9 station broadcasts and we want to receive it, we are not receiving the NTSC frequency for channel 9 (186-192 MHz). Digital channel 9 is broadcast in the UHF band. For example, KTVU Channel 2 digital TV is broadcast on channel 56 (722-728 MHz), not on channel 2 (54-60 MHz).

Why move up to UHF instead of using VHF? More losses with higher frequencies, but can use **smaller parts**. Less wind loading of the antenna, and can have a higher gain antenna for the same physical size.

Particular implementations of MPEG have made numerous additions to the basic program specific information to provide more sophisticated navigation tools, enable interactivity and give additional data and software to the user. The European DVB standard contains Service Information (SI), containing such things as a Bouquet Association Table (BAT), a Time and Date Table (TDT) and a Running Status Table (RST) as well as more detailed specifications of how conditional access data is transmitted. The American ATSC standard has equivalent but different extensions to the MPEG tables.

Europe modulates DTV using COFDM, United States modulates using 8-VSB.

The PSIP tables can also convey up to 16 days of programming information. Consumer receiver manufacturers can use PSIP data to display interactive program guides to aid navigation of channels in the DTV receiver. A tutorial on the subject is provided by the Sarnoff Corporation. TVs may not store all 16 days of information; Steve's TV only stores 14 days. Mistakes do happen, where schedule says one thing and the actual show is different.

The first application of PSIP you'll most likely see upon tuning a channel is a station identifier. It might tell you the station call letters and their channel number. Information provided is fully programmable by the TV station. TV stations have long identified themselves by their analog channel number. In their DTV allocations they might correctly identify themselves as NBC Channel 36, ABC Channel 53, or CBS Channel 60. Since the information provided is at the discretion of stations, they could set their PSIP to identify themselves by their analog channel number or what ever else they think is appropriate. The number in the PSIP doesn't even have to be a real channel number.

The PSIP also provides the number of program sources you might expect within that DTV channel (how many sources are multiplexed within the stream for that channel). Stations that do multicasting during the day and HD in the evening might leave the PSIP set for daytime activity, just not providing the additional programming indicated by the PSIP during the evening hours when high definition is available.

PSIP also provides program guide information. The view should be able to interact with this information. The DTV tuner might be programmed to use PSIP information for Direct Channel Change (DCC).

If a station is multicasting, providing several different programs at the same time, and the viewer can set a preference to see one type of programming as they enter the channel. This option might also be set to go to the program guide first instead of any of the actual programs.

DTV receivers attempt to draw upon the PSIP information in tuning channels. As stations change the channel number in their PSIP the DTV receiver will change the order in which they appear. Depending on the tuner's capability, if you want to tune a particular station you can enter their actual DTV channel number or their PSIP number. If you are using the channel up or down key to cycle through channels and you land on a PSIP indicating multiple channels available, yo might find yourself cycling through all of the allocations even if only one is active, before being able to go on to the next channel.

Is there an FCC requirement for broadcast stations to label their broadcasts digitally with their call signs? Steve hasn't heard of this requirement.

PSIPs can even redirect one channel to another, for example if there was a big announcement from the government, stations could redirect to the only station covering this.

Error Protection

The MPEG standard does not specify how error protection should be applied to the bitstream because this is specific to the transmission or recording medium being used. In broadcasting, a common approach is to add error protection to each 188-byte transport stream packet, for example a Reed-Solomon code.

Error protection is important because the variable-length coding, recursive prediction and limited resynchronization capabilities that are features of MPEG mean that single uncorrected errors in the video elementary stream can make a large mess.

Timing Information

Perhaps one of the most important features of the MPEG Systems Layer is that it provides a mechanism to enable the decoder to remain in step with the encoder when both are working in real time. The phase locked loop does this. The decoder has to be able to recreate the master clock that was used at the encoder, so that for every picture that enters the decoder, one leaves the decoder. It then has to know exactly when, according to this clock, it should remove data from the buffer for decoding and presentation.

The video-locked master clock at the encoder side drives a counter whose value is periodically multiplexed into the transport stream as the **Program Clock Reference** (**PCR**) and transmitted without passing through any of the large elementary stream encoder buffers. It is demultiplexed before passing through any large decoder buffers so that is has been subjected to a minimum and fixed overall delay. The demultiplexed value is compared with a locally generated count and the difference used to control a local phase-locked-loop clock generator. In this way the decoder clock is kept in sync with the encoder clock. Note that there is no need for the bit rate or channel clock to have any relation to the encoder or decoder clocks, or even to be constant. Different programs may have independent program clocks, or a common clock can be shared by several programs. A given station will use the same clock for all of its multiple subchannels.

Recreate the Master Clock

Diagram 1 and handout.

Time Stamp for Decoding

Why have a time stamp for decoding?

From its program clock the encoder knows what time it received each picture. It therefore knows what time the decoder should present each picture to the display. It calculates this time for each picture and transmits the information as a timestamp in the picture's PES header. The timestamp stays with its picture, passing through the encoder buffer. At the decoder it is compared with the local program clock and used to control the decoding and presentation of each picture.

MPEG-2 Video in Digital Broadcasting

Pre-processing

Any compression system aims to recover redundancy from the bitstream and encode what remains. Random noise on the source picture is by definition not redundant and will not be removed. The encoder will have to devote its bit rate resources to encoding the noise at the expense of the unpredicted parts of the picture. Noise reduction at the input to an encoder is therefore highly desirable. Bringing up gain brings up noise; properly lit scenes allow the gain to be kept at zero dB and minimize noise.

Other preprocessing may be worthwhile. At low bit rates, the balance between MPEG artifacts and overall lack of sharpness may sift to the extent that it is worth low-pass filtering and subsampling the input picture. Low-pass filter removes detail and some noise. It is also a good idea when starting with a composite signal is to apply a high-quality decoder which minimizes cross-color and cross-luminance, both of which are undesirable and eat up bit rate. The dot crawl is an example.

A good NTSC decoder can cost $16000.

Graph (in diagram 2) shows how good noise reducer improves efficiency of MPEG encoder. Two encoders working with fixed coding quality and variable bit-rate on a typically noisy source. Encoder 2 has no preprocessor while encoder 1 has preprocessing. Preprocessing can be worth 2 Mb/s per channel on typical material. Further improvements to MPEG-2 encoding quality can be achieved by analyzing the video signal and using the results of the analysis to control the encoder. For example, 50 Hz video signal derived from 24 Hz film using 3:2 pulldown can be analyzed so that the 3:2 phase information is passed to the encoder, which can then make use of the MPEG-2 syntax to avoid retransmitting the repeated field.

3:2 Pulldown

If have frames A, B, C, and D in a film strip, put on video the following: A1 & A1, A1 & B2, B1 & C2, C1 & C2, and D1 & D2.

A included three times, B included twice, C included three times, D included twice. Can refer to http://www.dvdfile.com/news/special_report/production_a_z/3_2_pulldown.htm for a graphic (at http://www.dvdfile.com/news/special_report/production_a_z/images/3_2_pulldown/3_2_a.gif) that illustrates this.

What about 25 Hz? Just speed up the film slightly (by 25/24 or 4%).

Concatenation

Many places in broadcast chain where necessary to decode an MPEG bitstream, perform some video processing, and re-encode result to MPEG. Why want to do this? Editing. When the processing substantial, involving spatial distortions and changes of scale, full-blown re-encoding is inevitable.

Why would this be bad? Anytime you convert, you may lose information. Re-encoding often working o signal maybe very close to decoded signal. Then, there is a significant concatenation loss because successive encoders may make different encoding decisions.

In the five or six generations typical in a broadcast chain, this loss can be substantial, as much as 6 dB in peak signal-to-noise ratio (PSNR). The concatenation loss can be reduced or even eliminated by reusing the coding decisions that were used in the previous encoder.

Where would we store the encoding decisions? This information is available in the bitstream and can be extracted from a suitably equipped decoder. International standards now exist for this re-coding data and for particular implementations such as history data and MOLE.

The MOLE system reduces the concatenation loss to zero by preserving all the picture-level and macroblock level information, and it offers great flexibility by carrying all the information in the decoded picture itself, in such a way that a suitable re-encoder can find out exactly where the re-coding data is valid and where the picture has been altered.

Concatenation is not the greatest term since it doesn't quite reflect the process of decoding and subsequent re-encoding.

Switching

Switching MPEG bitstreams, especially in real time, can be difficult because of the variable numbers of bits per picture and because of the need for access to previously decoded frames when forming motion compensated predictions. There are two main approaches to the problem.

Bitstream splicing relies on choosing suitable points in the bitstream where switching can be carried out. This is relatively simple to implement **but is limited to a straight cut** (no fades or fancy transitions), imposes severe restrictions on the choice of exact switching point and often imposes limitations on the original encoding process, forcing it to achieve a certain buffer occupancy at a certain time. Think of how GOPs are set up; one would need to transition between GOPs so that one doesn't start a cut in the middle of a GOP (not having the I frame).

The other approach is to decode the bitstreams and perform the switching function in the decoded video domain. Requires special decoders and encoders. This offers complete flexibility but can only be done without loss if re-coding data is used. The MOEL system allows a completely standard digital video mixer or DVE to be used while eliminating concatenation losses whenever this is possible.

Transcoding

Transcoding coves any process where one bitstream is converted to another. The conversion may be to another but related coding standard, such as DV, or it may be to another MPEG bitstream at a reduced rate. In both cases, it is important to make proper use of recoding data wherever possible. Sometimes the quality of transcoding can be significantly improved if further use is made of information about the previous encoding process. For example, the use of a maximum a posteriori probability (MAP) quantizer can lead to a transcoded picture quality that is as good as standalone coding at the transcoder's output bit rate.

Post-processing

The MPEG standard assumes that there may be some post-processing in the form of chrominance sample rate up conversion. The standard does not really specify exactly what should be done by way of post-processing for display. There are several ways in which decoders may differentiate themselves from competing devices by applying post-processing, for example:

- concealment of uncorrected errors – the MPEG standard provides for the transmission of concealment motion vectors which can be used in repairing or replacing corrupted blocks, but as usual does not specify how such information should be calculated or used
- reduction in visibility of blocking artifacts by adaptive filtering in response to some measure of the blockiness of the decoded picture (blocks have sharp edges, so are blurring the details of the blocks)
- nonstandard enhancements to MPEG-2 in which user information or metadata is used to convey instructions to the display processor on how to apply post-processing

Monitoring

Monitoring or measuring coded picture quality is a problem. Must take place on many levels, including bitstream integrity and consistency of systems-layer tables. At level of video elementary stream, there are two main approaches to measurement or monitoring, each of which has its place.

First is double-ended measurement, in which a decoded picture is compared with a source. The metrics used vary from the simple peak signal to noise ration (PSNR) to a number of methods that try to take the properties of the human visual system into account. Double-ended measurement useful as labtool but not for most monitoring purposes since no access to uncompressed source.

Other MPEG Flavors

MPEG-1

Precursor to MPEG-2. Designed to compress progressively scanned, half-resolution images to 1.5 Mbit/sec so that a feature film compresses onto standard CD. A subset of MPEG-2, lacking handling of interlaces sequences.

MPEG-2 Levels and Profiles

MPEG-2 decoder complies with standard if applies entire MPEG-2 gamut on all possible picture and sample rates likely to be encountered. The standard deals with this by defining a two-dimensional matrix of compliance points at which decoders can be specified.

One dimension is that of levels, defining different sets of picture size and clock rate limits. Low level handles MPEG-1 resolution. Main level for SDTV, High-1440 and High levels for HDTV resolutions.

Other dimension is profiles, different degrees of complexity. We mostly talk about Main profile.

Simple profile without B frames. Some scalable profiles more complex, having a base level that is higher priority than some higher levels, so that at least there is a base fallback in case can't present all the data.

MPEG-4

Differs from MPEG-2 by handling visual objects rather than just pictures. Diagram 3.

Video information coded separately as synthetic or natural objects, and resulting bitstreams multiplexed together. Decoder contains compositor putting decoded objects back together.

In coding natural video material, MPEG-4 provides some enhancements to the MPEG-2 toolkit, such as adaptive DC prediction, AC coefficient prediction, reversible VLC coding, global motion compensation, quarter-pixel motion estimation and shape-adaptive DCT coding as well as wavelet coding of textures and the use of sprites. But, people are finding ways to incorporate these things into MPEG-2.

However, consensus seems to be that the chief interest of MPEG-4 is in offering increased functionality rather than a huge leap in coding efficiency.