Transmission of High Bit Rate Studio Streams over IP Networks

1 Background

The Pro-MPEG Forum has already published guidelines in Code of Practice #2 to aid interoperability for the transmission of professional video over a variety of networks. The Operating Points described there have proven useful targets for equipment manufacturers and service providers, as well as providing a set of test-points for formal interoperability tests. Further information is available on the Pro-MPEG web site www.pro-mpeg.org.

In recent years it has become apparent that IP-based networks will become increasingly important for delivery of professional content. In the interests of interoperability, common approaches to the issues presented by such networks are desirable. End devices created by various manufacturers need to operate correctly with each other, and with networks using equipment from various vendors.

Pro-MPEG Forum Wide Area Networking has defined a suitable set of solutions to these issues for compressed video in Code of Practice #3 – "Transmission of Professional MPEG-2 Transport Streams over IP Networks". The approach taken by standards bodies and industry organisations has been included in the discussion leading to its publication. This Code of Practice addresses similar issues for High Bit Rate Studio Streams.

2 Scope

The application space being addressed is limited to Contribution and Primary Distribution applications.

This standard covers Uncompressed Standard Definition Video, running at 270Mbit/s in a way which will not prevent the carriage of compressed formats which use the same framing structure (e.g. SDTI). It is also intended that this document will be applicable to systems running at 360Mbit/s and HD rates to 1.485Gbit/s and possibly higher, but these are not the primary focus.

Where practicable this document re-uses practices laid down in Code of Practice #3.

To aid clarity to the code of practice the following words have been used:

- "shall" – mandatory to be compliant with this code of practice
- "should" – optional to be compliant but improves interoperability if adopted

2.1 Multicast support

Multicast needs to be supported both for a transmitting edge device sending to multiple endpoints, and for a receiving edge device to be able to receive a multicast transmission.
Detail on the multicast scenarios is given in section 5.

3 User Performance Requirements

The performance which is needed, and which must be provided by systems carrying professional video over IP is inextricably linked to the definition of professional video.

3.1 Baseline network requirements

In order for a system supporting this code of practice to function correctly it is necessary for the bandwidth available in the network to always meet or exceed that required by the IP stream generated by the system.
The total error rate on the network must be low enough to allow the error correction system referenced from section 4.4 to produce an output stream that can be successfully decoded.
The network jitter must be low enough to satisfy the requirements in section 4.6.
The network latency performance is described in detail in section 4.7.

3.2 System output error rate

The permissible system output error rate for professional applications is usually more stringent than that for domestic or industrial applications. The precise value is a topic for negotiation between customer (broadcaster) and network provider.
The FEC scheme proposed in the Code of Practice is configurable, so can support a range of un-corrected error rates across a range of network conditions.

3.3 System latency performance

Observations reveal that a number of professional applications have demanding round-trip delay requirements from the entire video system (from camera input to monitor output for both directions) . A round-trip delay of 400ms (reference if possible) is widely accepted as the worst that is easily usable for live interview applications, without special training. Camera control by remote telemetry has been shown to require a round trip delay as low as 200ms (reference if possible). Care needs to be taken in the level of FEC applied to ensure that there is not a detrimental effect on the system's overall latency.

4 Transmission Protocols

4.1 IP carriage issues

The size of the output IP packet from a transmitting device shall be limited so that IP fragmentation does not occur at the output of the device. The IP 'don't fragment' bit shall be set. As end-point devices will typically be connected to Ethernet style networks, this limits the maximum transmission unit (MTU) to 1500 bytes. The MTU on links between intermediate nodes in the network may be lower than this, so care shall be taken to ensure that IP packets are not lost due to the 'don't fragment' requirement. Because of this MTU limitation it will be necessary to always split the data from one video line across multiple IP packets. Some IP networks prioritise different packet sizes differently, so it may be desirable to split the video data in a way that gives packets of approximately the same size.

4.2 RTP/UDP/IP Mapping

The use of RTP is required, as it provides a standard header for the packets.
The RTP format proposed is that from RFC3497 with the following extensions:

- For Standard Definition, the RTP timestamps shall be derived from a 27MHz clock locked to the input, not from the 148.5MHz clock required by RFC3497.
- For other video systems the RTP timestamp shall be the word clock rate of the video system.
- To assist with picture concealment in the event of lost packets, the data in a packet shall only come from one line[1] of video. As an implication of this the EAV field will always occur at the start of a packet.
- In addition to this, there shall be a custom RTP header extension to assist with defining where in a line a particular packet starts which equipment which complies with this code of practice shall adopt.

[1] For the definition of 'a line', see section 2 of RFC3497. Though this is an HD reference, it is also valid for the Standard Definition case

The full RTP packet is thus:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| V |P|X|   CC  |M|    PT       |      sequence# (low bits)     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     time stamp                                |   A
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        ssrc                                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| header extension type = 0000h |header extension length = 0001h|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+   B
|       reserved        |             line byte offset          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    sequence# (high bits)     |F|V|         line no            |   C
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
:                      HBRSS data                               :
:                                                               :
|                                                               |
+---------------------------------------------------------------+
```

In the above diagram the red section A is the RTP (RFC3550) header, the green section B is the RTP header extension (formatted as specified in RFC3550), and the blue section C is the payload (RFC3497) header.

The fields in the above are defined as follows:

V	The RTP version number = 2, as defined in RFC3550
P	Padding bit, shall be 0, as there is no means of detecting the end of the valid data in the packet if there was padding present.
X	The RTP header extension bit, shall be set to 1
CC	The CSRC count, as per RFC3550, this field should be set to 0 (no CSRC records) for simplicity.
M	The marker bit, shall be set to 1 to mark the last packet of the video frame, and to zero for all other packets (as defined in RFC3497)
PT	The payload type field. See section 4.3
Sequence#	This is split into the low bits, which shall be inserted into the RFC3550 location, and the high bits which shall be inserted as per RFC3497.
Timestamp	The RTP timestamp, as defined above.
Header extension type	As this is the first header extension, the type field shall be zero.
Header extension length	Shall be set to 0001h to indicate that the header extension is a single 32 bit field.
Reserved	A 12 bit field which shall be reserved for future uses, possibly for a packet length field to allow the use of padding.
Line byte offset	This shall be defined to be a byte count along the video line of the first byte of the payload, starting with zero at the EAV field for a line.

F	As defined in the video timing signals. F=1 identifies field 2 and F=0 identifies field 1 (from RFC3497)
V	Also defined in the video timing signals, V=1 during vertical blanking, and V=0 elsewhere in the picture (from RFC3497)
Line no	The line number of the video encapsulated, as defined in the video format. RFC3497 limits this field to 12 bits, which only supports up to 4096 lines. The RFC does however have two 'Z' (zero) bits which shall be used to extended the line number range to 16k lines. For video formats with 4096 lines or less the stream will still be RFC3497 compliant.

For interoperability, it is required that equipment shall only depend on the main RTP communication between sending and receiving units, and shall not require any additional communication, including but not limited to RTCP.

If equipment manufacturers wish to make use of such additional communication to improve the operation of their units then they are not forbidden from adding this, provided that the equipment can function correctly in its absence.

4.3 RTP Payload type definitions

As RFC3497 requires the use of a dynamic payload type, and there is currently no recommendation for an automated signalling system it is required to statically assign the dynamic payload types, as was already required for the FEC stream in CoP#3. As the first dynamic payload type (96 decimal) has already been used for the FEC stream, the following allocations shall be used:

97: Standard Definition SDI and other formats using a 27MHz word clock
98: Formats requiring a 36MHz word clock
99: High Definition formats requiring a 148.5MHz word clock.

Future allocations may be made if other video formats are required. The dynamic payload type range allows values up to 127 to be used.

4.4 FEC Scheme

The FEC scheme shall be as defined in section 4.4 of Code of Practice #3.

Due to the much higher rates for HBRSS streams, the number of packets lost for a given duration outage will be far more than for the Transport Stream case.

As a result, the recommended values for L and D change, the new suggested limitations should be:

L * D <= 1500
1 <= L <= 255
4 <= D <= 20

Note that at 270Mbps the maximum size FEC matrix will cause a latency of over 70ms. As the requirement for using HBRSS streams are often driven by latency demands, care must be taken to ensure that the FEC latency does not compromise such requirements.

4.5 Timing recovery

The RTP timestamp defined in section 4.2 is sufficient for recovering the timing of the HBRSS stream.

Because of the standard definition of an RTP timestamp, the timestamp will be the target output time of the first byte of the payload.

The precise method of recovering the timing for the output is implementation specific. There are several approaches, with varying advantages and disadvantages.

For many applications it will be desirable to have an external synchronisation input into the receiving unit, so that its output is locked to the other feeds being used in the same studio.

4.6 Jitter Tolerance

Network jitter can be absorbed by buffering at the receiver. There are two components in a typical IP network jitter issue. There is a first high-frequency component, caused by load spikes in the network. These tend to be quite small in value, of the order of ±10-15ms. On networks carrying Internet or other data traffic there is a 'wander/drift' component, as the loading of the network varies over a 24-hour period. This will typically be larger, of at least ±30-40ms. For the benefit of simplicity, these can be treated as one by providing a 'jitter budget' buffer of 120ms. This buffer should be run half full on average, providing a 60ms latency. For flexibility, it is recommended that system designs should make it possible to modify the size of this buffer, as networks can have either significantly better or worse jitter performance.

The jitter absorption needs to be handled carefully, to ensure that the output signal is compliant to the output format timing requirements.

4.7 Latency

Latency within an IP adaptation unit is bounded by a combination of the jitter tolerance buffer, the FEC system and any additional buffering required as part the clock recovery mechanism (not covered by this recommendation).

There are additional latencies caused by the IP network transmission.

A typical FEC latency is given by the formula: (L*D*packet-length*8)/ip-stream-rate

For example, with L=10, D=10 and an approximate packet length of 1100 bytes, for a 270Mbit/s video stream this formula gives a typical FEC latency of 3ms.

4.8 Re-order Tolerance

Packets travelling over IP networks are not guaranteed to arrive in the order sent. Sequence numbering is provided by RTP, which should allow this effect to be corrected within the receiving end equipment. Any re-ordering that is present is likely to be of a small order, less than 10 packets out of place.

The FEC system will correct the levels of packet re-ordering likely to be encountered. If a packet is grossly out of order then it is discarded, and will be corrected by the FEC scheme as a lost packet when it would have been expected. If a system holds multiple FEC matrices then it can tolerate re-ordering within this group of matrices. Systems should provide a means to be able to handle packets which are only out of order by a very small amount, but which occur the wrong side of their FEC matrix boundary as a result.

4.9 Encryption

Encryption should be handled either on the raw video data before encapsulation, or on the IP data after encapsulation and so is outside the scope of this document.

5 Signalling Protocols

5.1 Network QoS protocols (RSVP/MPLS/DiffServ etc)

The terminal equipment should allow the IP header TOS byte to be fully configured. This allows both traditional TOS values, and DiffServ markings.
Bandwidth reservation shall be handled using out of band mechanisms - the end devices are not required to support a bandwidth reservation protocol.

5.2 Session control protocols (SAP/SIP/SDP/RTSP/RTCP etc)

Further work is required to define the signalling systems to be used, and the parameters for these systems.

As such the initial baseline is that systems can have the required parameters manually configured. The following parameters list is the required minimum:

- As the system is UDP based, and there is no commonly accepted standard UDP port number or range for RTP systems, it is recommended that the sender should allow the destination port number and IP address (which may be a multicast group) to be configured. There is an RTP requirement that any port number chosen shall be an even number.
- It is required that the FEC on a sending device can be enabled or disabled - though it is recommended that FEC should be used for all contribution applications. There may be some applications where broadcasters are trying to maximise the use of the bandwidth, and are happy to accept the extra errors in the signal in exchange for the extra bandwidth available.
- The receiver shall be able to cope if no FEC stream is received, because the sender may or may not be transmitting one.
- For configuration simplicity, where FEC is present, it shall be sent on the UDP port number two higher than the main data port number configured (The port number one higher is reserved for use by RTCP).
- To support multicast, there should be an option on receiving units to allow a multicast group address to be configured.
- The receiver should use IGMP version 3 to join and leave multicast groups when required. As some network equipment may not yet support IGMP version 3, there may be applications where earlier versions of IGMP are required.
- The port number the receiving unit will listen on should also be configurable. Again, there is a requirement that the RTP port number must be even.
- As mentioned earlier, the value for the IP TOS byte should be configurable.

6 Management Protocol

As applications mature there will be a desire to adopt a common management interface, this requires further study.

-- END --

