INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 MPEG2014/M35087
October 2014, Strasbourg, France

Source **Telecom ParisTech**
Status **For consideration at the 110th MPEG Meeting**
Title **Clarifications on ISOBMFF negative composition offsets**
Author Cyril Concolato, Jean Le Feuvre

1 Introduction

Negative composition offsets are allowed in CTTS box for which version 1 is used. In this case, the effective composition time can be found using two ways:

- by checking the CompositionToDecode box in the track, and add cslg. `compositionToDTSShift` : this is the safest approach, as the file creator explicitly gave the missing offset for reconstruction
- or by analyzing the CTS offsets in the ctts table or in the trun entries, and consider use the smallest value found.

This contribution reviews some issues when processing CTTS version 1 streams.

2 Sample storage

2.1 Single frame inversion

Consider the following video structure:

Frame	DTS	CTS	Ctts v=0	Ctts v=1 Shift= 1	CTS cslg omitted
I_0	0	1	1	0	0
P_3	1	4	3	2	3
B_1	2	2	0	-1	1 (DTS > CTS)
B_2	3	3	0	-1	2 (DTS > CTS)
P_6	4	7	3	2	6
B_4	5	5	0	-1	4 (DTS > CTS)
B_5	6	6	0	-1	5 (DTS > CTS)

In order to synchronize with the audio track, the video will have an edit list with one entry lasting for the movie duration, 0 as a movie start time and 1 as a media time: “At movie time 0 play from the I frame”.
In this case, when the cslg is omitted, the CTS reconstructed is broken. A file reader needs to check the smallest CTTS offset to adjust the CTS. Assuming it doesn’t do so (since such a

rule/guideline is nowhere to be found), it will find for CTS=1 the frame B_1 instead of the frame I_0, and one frame will be skipped!

We therefore conclude that the file reader SHALL estimate the min CTS shift (smallest entry in the CTSS v1) to compensate. This should be written somewhere in the specification.

2.2 Single then dual frame inversion

Consider the following video structure (GOPs are shortened for simplicity !). At the beginning of the stream, the encoder uses only one frame inversion but plans to switch to 2 frame inversions in the short term. Each gop is carried in a single segment/fragment, and the inversion of an extra B happens at the beginning of the second segment

Frame	DTS	CTS	Ctts v=0	Ctts v=1 Shift= 2	CTS cslg omitted	CTS cslg recomputed shift=1
I_0	0	2	2	0	0	1
P_3	1	5	4	2	3	4
B_1	2	3	1	-1	1 (DTS > CTS)	2
B_2	3	4	1	-1	2 (DTS > CTS)	3
P_6	4	8	4	2	6	7
B_4	5	6	1	-1	4 (DTS > CTS)	5
B_5	6	7	1	-1	5 (DTS > CTS)	6

Frame	DTS	CTS	Ctts v=0	Ctts v=1 Shift= 2	CTS cslg omitted	CTS cslg recomputed shift=2
I_7	7	9	2	0	0	9
P_{10}	8	12	4	2	3	12
B_9	9	11	2	0	1 (DTS > CTS)	11
B_8	10	10	0	-2	2 (DTS > CTS)	10
P_{13}	11	15	4	2	6	15
B_{12}	12	14	2	0	4 (DTS > CTS)	14
B_{11}	13	13	0	-2	5 (DTS > CTS)	13

As can be seen, the timeline initialization of the first segment is broken, even when the client estimates the CTStoDTS shift. There will be a frame freeze of one frame duration between the two segments, as the first segments ends at CTS=7 and the second starts at CTS=9

Furthermore, if an edit list is present, the first frame presented by the client will not be the frame intended by the author, since there is a one-frame mismatch in timing.

We can now conclude that there is no guarantee that the CTStoDTS shift computation at the client side will be correct.

3 Conclusion and proposal

We believe that the CSLG box should be made mandatory when CTTS and/or trun are used in version 1, and that a COR should be issued to fix this problem.