108th MPEG Valencia, Spain, 31 Mar. - 4 Apr. 2014, Meeting Report

Panos Kudumakis
qMedia, Queen Mary University of London

Contents

1 Publish/Subscribe Application Format (PSAF)

Publish/Subscribe (PubSub) is an established communication paradigm where senders do not communicate information directly to intended receivers but rely instead on a service that mediates the relationship between senders and receivers. In PubSub senders (called Publishers) post information on and receivers (called Subscribers) declare their interest in a certain type of information – before or after a publication – to a service.

A typical workflow of a content distribution context that can benefit from the PubSub modality is given by *Table 1* where the need for 4 information formats is highlighted in italic.

Table 1 – Steps in multimedia Publish/Subscribe

	Step	**Information type required**
1	Creator stores resource	Resource ID or locator
2	Creator stores information on resource	*Resource format*
3	Publisher publishes information on resource	*Publication format*
4	Subscriber subscribes to a class of resources	*Subscription format*
5	Service matches subscription with publication	No information type
6	Service issues notification(s)	*Notification format*
7	Subscriber opens notification, requests/plays resource	No information type

This document collects the requirements to be satisfied by the payloads utilised by a Publish/Subscribe mechanism for media applications. The definitions of *Table 2* apply.

Table 2 – Users in multimedia Publish/Subscribe

Creator	Creates a Resource
Publisher	Sends Publications to MSPs
Subscriber	Sends Subscriptions to MSPs
Match Service Provider (MSP)	Performs matches of Subscriptions with Publications on the request of Pub-Sub users
Pub-Sub user	Is a Publisher or a Subscriber

The walkthrough is graphically depicted in Figure 1.

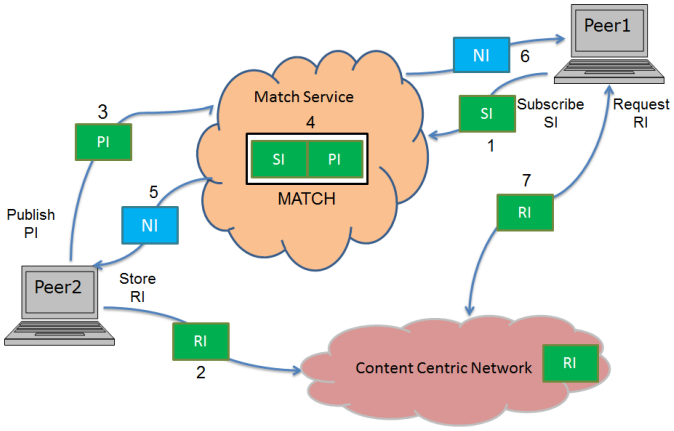

Figure 1: Implementation of Publish/Subscribe for multimedia applications

Requirements

Pub-Sub users of Publish/Subscribe for multimedia applications shall be able to

1. Define users to be/not to be notified of an event in the form of
    a. Enumerations
    b. Groups of users
    c. Conditions that users should satisfy
2. Reference the Resource via a standard information package containing at least the following information elements
    a. Descriptions of Resource
    b. Rights and conditions for use of Resource
    c. List of users to be notified that a specific use of a Resource has been made
3. Select Match Service Provider(s)
4. Prescribe which Publications shall be/shall not be considered in computing matches with Subscriptions and which or Subscriptions shall be/shall not be considered in computing matches with Publications
5. Prescribe to Match Service Provider users to be/not to be notified of an event in the event of a match
6. Define validity of Publication/Subscription (i.e. the period of time defined by a start and end time the Match Service Provider shall notify matches)
7. Guarantee authenticity of Resource, Publication, Subscription and Notification Information
8. Update Publications/Subscriptions

Technologies

The following MPEG technologies are used to specify the PSAF

1. Digital Item
2. Digital Item Identifier
3. Simple Metadata Profile
4. MPEG Query Format
5. Rights Expression Language
6. Event Reporting

Outside MPEG the W3C Signature Recommendation.

Output documents

N14521 - Requirements for Publish/Subscribe Application Format (PSAF)

N14381 - Preliminary WD of Publish/Subscribe Application Format

2 Candidate Technology for Dynamic Range Control

An additional tool has been proposed at the 108th MPEG meeting in Valéncia in [M33147] to reduce the peak bitrate for MPEG-D Dynamic Range Control [N14261]. It has been the consensus of the group to perform a cross-check of the results presented in [M33147]. In case of a positive cross-check result, it has been the

consensus of the Audio Subgroup that the modifications proposed in [M33147] will be accepted and included in the CD of MPEG-D Dynamic Range Control (N14455, Text of ISO/IEC 23003-4 CD, Dynamic Range Control).

Summary of Proposed Technology

In [M33149] optional syntax elements in MPEG-D DRC has been proposed, which allow for an adaptive distribution of DRC gain payloads over several frames. It is referred to as *payload splitter* in the following. The proposed mechanism is helpful for applications where high quality DRC encoding is desired but strongly varying peak bitrates are critical at the same time. The distribution of the DRC gain payloads can be controlled in a flexible way dependent on the allowed DRC encoding delay. The distribution of DRC gain payloads has no effect on the output gain sequence.

Planned Work

The performance of the *payload splitter* will be further investigated in additional experiments with focus on the context of MPEG-H 3D Audio, e.g., by including a larger number of gain sequences. The DRC gains sequences (for DRC-1/2/3 [M33151]) will be generated for use cases that are considered relevant for MPEG-H 3D Audio.

The use cases may include

- Movie sound with different language tracks for the dialog.
- Adaptation to specific playback configurations (e.g. downmix or target level).
- Adaptation of the audio to specific listening environments ("noisy", "late night", …).

Alternatively to the above use cases, multiple DRC gain sequences may be randomly chosen from the current DRC test set to simulate transmission of multiple independent DRC sets. The obtained bitrate will not only be presented separately, but bitrate differences will also be given relative to the overall bitrate including audio bitrates that are typical for MPEG-H 3D Audio content. The results will be reported as an input contribution to the 109th MPEG Meeting.

References

1. ISO/IEC JTC1/SC29/WG11 N14261, "*WD1 Text of Dynamic Range Control*", January 2014, San Jose, US
2. ISO/IEC JTC1/SC29/WG11 M33149, "*Technical Description of a Tool for DRC Technology*", April 2014, Valencia, Spain
3. ISO/IEC JTC1/SC29/WG11 M33147, "*Core Experiment on Improving MPEG-D DRC Technology*", April 2014, Valencia, Spain
4. ISO/IEC JTC1/SC29/WG11 M33151, "*Proposed Text on DRC and Loudness Technology in MPEG-H 3D Audio*", April 2014, Valencia, Spain

Output documents

N14455 - Text of ISO/IEC 23003-4 CD, Dynamic Range Control

N14456 - WD of CD Software for Dynamic Range Control

N14457 - Candidate Technologies for Dynamic Range Control

3 Issues related to MPEG-H 3D Audio

Immersive Audio Rendering

At the 108th MPEG meeting in Valencia, "Crosscheck Report on Immersive Audio Rendering" [1] was submitted and introduced. The basic idea of the contribution is to provide an Immersive Audio experience to the majority of consumers with reproduction systems over legacy layouts. After the discussion of the results, the Audio Group concluded to perform another crosscheck.

Crosscheck Timeline

The Immersive Audio will be compared with RM2 rendered signal (baseline). The following *Table 3* outlines the planned steps to investigate the proposed rendering.

Table 3 - Crosscheck Timeline

Activity	Completion Date	Responsibility
Send test materials	May 14th, 2013	Samsung
Cross check of the Samsung's Immersive Audio Rendering	May 28th, 2013	Samsung, FhG IIS, ETRI
Report of cross check results and updated CD Text	At Adhoc Meeting	Samsung
Approve by the Audio Group	At Adhoc Meeting	Audio Subgroup
Release the reference software	Next Meeting	Samsung

References

1. MPEG document M33138, "Crosscheck report on Immersive Audio Rendering," 108th MPEG Meeting, March 2014, Valencia, USA.
2. MPEG document N13411, "Call for Proposals for 3D Audio," 103rd MPEG Meeting, January 2013, Geneva, Switzerland.

Evaluation of BRIR Parameterization

In the 108th meeting Huawei proposed a new metric (Differential Energy Decay Curve, DEDC), to estimate mixing time in user BRIR databases. This solution seems interesting for an accurate, consistent and universal estimation of mixing time in both QMF and Time domain. The present Experiment aims at formally testing the effects of this metric in TD and FD binauralization. The experiment will be run at least on 1) two test sets (Short, Medium, Long) used for binaural CE and it can be done on 2) additional test sets will be added which will be provided by Huawei and agreed by the crosscheck laboratories ETRI/Yonsei/WILUS, ORANGE).

Complexity of the algorithms using the original and the new metric will be taken as one evaluation parameter, together with listening tests results.

Listening tests

Test sites are HUAWEI, ETRI/Yonsei/WILUS, and Orange. The test design definition should make a consensus between HUAWEI, ETRI/Yonsei/WILUS, and Orange.

Activity	Completion Date	Responsibility
Test design definition	Apr 18, 2014	HUAWEI, ETRI/Yonsei/WILUS, Orange
Provide Test BRIRs,	Mai 16, 2014	HUAWEI
Make available binauralized waveforms and parameter coefficient files for each BRIR	Jun 9, 2014	HUAWEI, ETRI/Yonsei/WILUS, Orange
Listening tests start	Jun 9, 2014	Test sites
Listening test reports	109th meeting	Test sites

Decoder Delay Alignment

The issue of compensation of non-predictable decoder delay was brought to the attention of the group in contribution M33231, It covers MPEG-H 3D audio codec behaviour at reconfiguration events. It was the consensus of the Audio Subgroup to investigate further and in particular quantify the delay component as well as the delays of the varying processing tools in the MPEG-H 3D audio decoder and to study which tools may require delay compensation.

Harmonization of interface data structures

The audio group recognizes that different data structures have been used, e.g. for describing loudspeaker layouts to describe such layouts in encoding (reference loudspeaker layout), loudspeaker decoding/rendering

(local speaker layout) and binaural rendering. These data structures should be unified as far as possible and meaningful.

Binaural Interface

The output document N14463 "Normative Interface for Binaural Renderer" defines a binaural interface. This document will be further reviewed for inclusion in the CD.

Output documents

N14458 - Request for Subdivision ISO/IEC 23008-3 CD

N14459 - Text of ISO/IEC 23008-3/CD, 3D audio

N14460 - WD of ISO/IEC 23008-3:20XX/Amd 1, 3D Audio File Format Support

N14461 - Listening Test Logistics for 3D Audio Phase 2

N14462 - Active Downmix Control

N14463 - Interface for Binaural Data

N14464 - Candidate Technologies for 3D Audio

N14465 - Workplan on 3D Audio

4 Exploration - Audio Synchronization Use Case

The audio synchronization scheme is envisaged for "second screen" applications where the 2nd screen content is synchronized to the 1st screen content. In this scenario, no common clock covering the 1st and 2nd screen devices is required, nor way to exchange time-stamps between the devices. Synchronization of the contents between the devices is done by using audio features (fingerprint) extracted from the 1st screen content.

For example, 1st screen content is distributed over existing broadcast system, and 2nd screen content is distributed over IP network, where 1st screen content is a regular TV program of car racing, and 2nd screen content is a drivers view of the car racing. Feature of the 1st screen audio content is used for the synchronization. The audio feature (fingerprint) stream of the 1st screen content is sent to 2nd screen together with 2nd screen audio/video content over the IP network. In the 2nd screen device, audio of the 1st screen content is captured by a microphone and its feature is extracted. The extracted feature from the microphone input and received feature from IP network is compared and time difference is computed. This time difference is used to align 2nd screen audio/video content to the 1st screen content. One of the greatest benefit of this approach is that transmitter/receiver system of main media stream (for 1st screen) need not be modified.

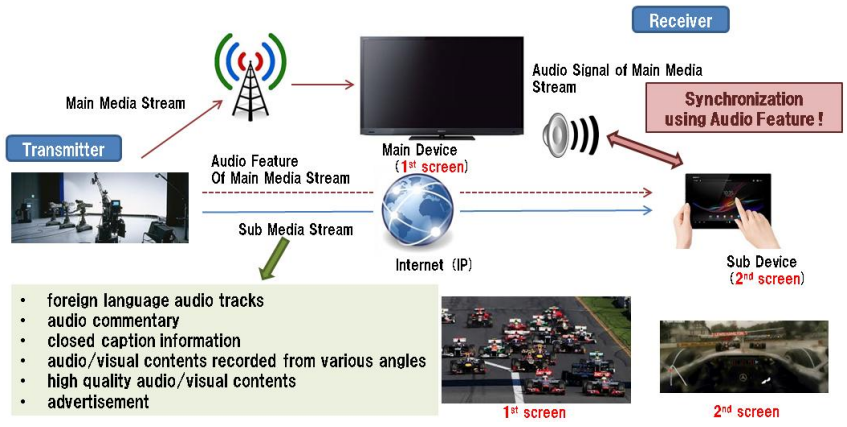

Figure 2: Audio synchronization use case of "second screen" application

The audio synchronization scheme utilizes Audio Object Type (AOT) "Audio Sync" so that audio feature stream for synchronization can be transmitted as an Elementary Stream (ES) with Time Stamp in MPEG-2 TS or MPEG-4 system's way. Then the Time Stamp of the Audio Sync stream can be used as a reference clock for the 2nd screen audio/video stream.

Status of the proposed synchronization scheme

The cross check results of the synchronization scheme using the MP4 File Format were reported in the 108th meeting, and the Audio Subgroup reached the consensus that the synchronization scheme be accepted as a part of the MPEG standard conditionally, that means:

- Audio Object Type (AOT=46) is defined for the audio feature stream in the MPEG-4 standard.
- The proposed audio feature extraction algorithm is accepted as a normative part of the MPEG-4 standard.
- The proposed audio feature and its syntax are accepted as a normative part of the MPEG-4 standard.
- The proposed feature *matching algorithm is accepted as an informative* part of the MPEG-4 standard.

References

1. MPEG document M31368, Proposed scheme for synchronization of multiple audio streams
2. MPEG document M32189, Cross-check report of audio synchronization
3. MPEG document M32236, Cross Check Report of Audio Synchronization Experiment
4. MPEG document N14150, Proposed Exploration of "Uniform Signaling for Timeline Alignment"
5. MPEG document N14270, Workplan for audio synchronization
6. MPEG document M33079, Cross-check report of audio synchronization
7. MPEG document M33161, Synchronization scheme using audio fingerprint for various use cases

Output documents

N14467 - Status of Audio Synchronization

5 Exploration - Storage of traceable media signatures

As ISOBMFF (14496-12) in combination with Common Encryption (23001-7) becomes increasingly utilized for the distribution of high-value content, there is a growing need to support coding tools to aid with the identification of the player model from which content has been copied without authorization. A common identification capability is known generically as "forensic marking".

A forensic marking system utilizes "variances" in the sample data that in aggregate are *unique to the playback environment*. Applying the forensic mark at the consumer player necessarily requires a secure environment such as that offered by modern DRM systems and/or "hardened" players. It is important to the process to guarantee that forensic marks are applied as intended during playback of the content or that tampering either results in unacceptable playback quality or cessation of playback – concepts which are at the core of modern DRM systems.

Specific forensic mark systems (out of scope of this study) use variant data to modify the sample data in ways that are not perceptible to the consumer, but result in sufficient changes to be detected by companion forensic mark decoders. Variant algorithms typically operate on the video samples, but can also apply to other sample types such as audio. The exact variant data that is utilized and the granularity of detection is dependent on the application and the specific variant algorithm used.

Forensic marking enables efficient technical content protection response to the discovery of unauthorized copies to more promptly prevent future occurrences. The ability to better respond to such events encourages publishers to continue to release high value content using ISOBMFF.

In fact the US movie industry has developed requirements for next generation content and its protection that are captured in some specifications from MovieLabs: Next Generation Video & Enhanced Content Protection.

Having a *standardized* variant sample framework for ISOBMFF enables:

- industry adoption of "forensic marking" support with a non-proprietary framework for better interoperability;
- support of ISOBMFF standards by services which require "forensic marking" ;
- increased availability of tools which support "forensic marking"; and
- enabled partial processing of variant data in silicon.

There is today no standardized way to deploy "forensic marking" with ISOBMFF files. Every proprietary system is different.

Note: This is *not* about standardizing any specific forensic marking technique or require any specific DRM System.

Output documents
N14518 - Draft Requirements for a Framework for sample variants in ISOBMFF

6 Exploration - Use Cases for Processing and Sharing of Media under User Control

In the actual use of MPEG technologies, there are many situations that require the media of its components to be private, with processing sharing under user control. Examples of this are:

1. searching an encrypted audio visual database with an encrypted query
2. identifying a spoken keyword in a private conversation, e.g. encrypted audio
3. removal of identification clues from multimedia content such as media, audio or speech
4. sharing multimedia content in a limited context, e.g. making a picture available to a limited list of persons, for a limited time, or for a specified purpose

MPEG has been developing very successful standards that process audio and video information based on its significant expertise in this domain MPEG is now identifying application domains and extract requirements to achieve both application of existing standards, and use of its expertise to develop new standards, supporting the private processing and sharing of media.

Use Case: Privacy-Preserving Media Search & Analysis
In this use case, Alice would like to search a media database that is owned by Bob. Alice encrypts her query so that what she is looking for is kept private and not revealed to an un-trusted party. The media in the database is also encrypted to keep it secure. A secure protocol is executed between Alice and Bob to perform matching without decrypting either the query or the media stored in the database. The best matching results are then returned to Alice (or a third party) without revealing the query to Bob, or the media in the database to Alice.

Rather than searching for media, it is also possible to consider a privacy-preserving analysis of the media. For example, consider the analysis of a speech signal from a conversation. The speech signal is encrypted to keep the conversation private. The encrypted speech signal is then sent to an un-trusted server for analysis, where keywords from the speech could be detected using a secure protocol.

The general problem of finding nearest neighbours in a privacy-preserving manner is applicable to a broad range of problems, many of which involve media such as images, video and audio signals. This problem can be broken down into two steps. The first step requires a method for privately computing distances, and the second step requires a method for privately finding the minimum distance to determine the nearest neighbour.

While the above example scenarios refer to the use of cryptographic primitives, there also exist methods to perform these operations using a form of secure fingerprints.

Use Case: Context aware multimedia privacy protection in video surveillance and social networks
In this use case, recorded video surveillance footage, or social networks share content that may become available to a large audience. Sharing of such content and the context of sharing such information is an essential issue.

The context could be a select group of people, purpose of sharing, time and date, metadata and the likes. Applications on smart devices tend to interact with platforms mutually without informing to the user the privacy and security concerns they may have.

In video surveillance, a number of consortia have been exploring various implications and problems that sharing of video surveillance footage can arise from ethical as well as legal points of views. Examples include EC funded network of excellence VideoSense (http://www.videosense.eu).

Likewise, several examples of applications and services exist that allow sharing of video in social networks, such as Socialcam, a mobile application which allows users to share with their friends as well as with public, various video.

Currently, the usage control of such content is either trivial, or complex, and inefficient. For example, Spotify and Yahoo News automatically publish songs and news one has listened to or read on the profile of users.

Privacy is not static nor deterministic but rather a dynamic and stochastic parameter which needs to take into account both the context in which the protection took place and the context in which a user accesses content and information. It is therefore important to devise a framework in any solution for user control, which explicitly takes into account not only the relationship between content and privacy, but also the context in which this relationship occurs.

Use Case: Privacy-Preserving Video Transcoding
For applications like dynamic adaptive streaming and screen content sharing, video content generated at its source needs to be transcoded to fit a large variety of client devices and dynamic network environments. When video contains privacy sensitive information and content and transcoding is conducted by an un-trusted party, it is necessary to keep privacy undisclosed while allowing transcoding transformations.

One possibility is to encrypt video content, or a segment (for a period of time, e.g., a particular scene) or a portion of it (for a region in the content, e.g., a person's face), in a transcoding friendly manner. This requires a good understanding of common transcoding transformations used in practice and what encryption schemes can be used to permit conducting these transformations while video content is encrypted using any of those schemes.

Output documents
N14514 - Use Cases for Processing and Sharing of Media under User Control

7 Exploration – Uniform Signalling for timeline alignment

MPEG has started an exploration activity on media stream synchronization in heterogeneous delivery environments, transported using existing MPEG systems technologies. The Systems subgroup is planning a seminar at the 110th MPEG meeting in Strasbourg to advertise this activity and collect more requirements/needs from the industry.

In a number of scenarios, ancillary timed content can be made available to enhance the experience of consuming some primary content. Examples include accessibility, sub-titling or captions, different audio tracks but also content that enhances the user's experience in other ways.

To enable a wide range of interoperable services, it is desirable that the availability and the alignment of ancillary media for the primary media are signalled using standard techniques and terms, in a uniform way, independently from formats and transport protocols.

MPEG has started an exploration activity on discovery and synchronization of auxiliary media streams for its Systems technologies (MPEG-2 Transport Stream, ISO Base Media File Format, MPEG-DASH, MPEG Media Transport). This exploration focuses on the problem of playing one content that is time-aligned to some other content, and possibly delivered using a different transport, physical network and/or encapsulated in different containers. Both contents may be played on the same device, or on different devices sharing synchronization information.

There are several use cases where it may be interesting for a service provider to deliver part of its content on one transport channel, for example free-to-air terrestrial or satellite broadcast, and part of it on another such as broadband IP connection. In order to depict a complete overview of the use cases they can be categorized and distinguished against several characteristics. Here is a non-exhaustive list of such criteria:

- Number of different contents
- Type of the content, e.g., audio, video, etc.
- Format of the content
- Number of devices playing-out at least one of the content
- Physical locations of the devices
- Physical networks the devices are connected to
- Requirement on synchronization accuracy (e.g., frame accurate or not)

Proposed Topics:

- Hybrid Broadcast - Broadband distribution for UHD deployments: a use cases perspective
- Hybrid Broadcast - Internet applications on companion screens
- HEVC and Layered HEVC for UHD deployments (potentially HDR ?)
- Inter Destination Media Synchronization – Social TV use cases
- Audio Fingerprinting based Synchronization

Round table discussion on issues on UHD deployment, second-screen apps, hybrid broadcast-broadband, HEVC, MPEG systems, ..., and demos.

The Systems group welcomes feedback on the above topics and suggestions of demonstrations, which are deemed of interest for MPEG activities.

Output documents
N14365 - Uniform Timeline Alignment
N14523 - Plan of Seminar on Hybrid Delivery at the 110th MPEG Meeting