SIViP
DOI 10.1007/s11760-009-0107-6

ORIGINAL PAPER

Modeling and delivering heterogeneous audiovisual content for group consumption

Kostas Karpouzis · Ilias Maglogiannis

Received: 25 September 2008 / Revised: 20 January 2009 / Accepted: 21 January 2009

Abstract The abundance of broadcast material, especially when it becomes available from a variety of content providers, makes the choice of a program genre and adaptation of presentation options to the preferences of a user, a welcome feature of the modern-day TV viewing experience. From a technical point of view, assembling and transmitting such heterogeneous content is in itself a daunting task, especially when intellectual property rights issues should be tackled. In addition to this, while there are a lot of options for filtering content with respect to the preferences of a single user, the common or aggregated choice of a group is hardly ever taken into account. Considering the fact that TV viewing and multimedia consumption in general are essentially a social activity, systems which package, filter and rank the available content or propose similar content to what is currently viewed should also integrate mechanisms to model group dynamics. This article presents an integrated, end-to-end architecture which assembles multimedia material, respecting the IPR of the content provider, and delivers it to a client-side mechanism which considers the preferences of all the viewers currently watching to filter and rank the available programs. In order to respect the established methods of producing content, this system utilizes concepts from adopted standards (MPEG7, MPEG21) to model processes and represents data and relations between the different entities of the system.

K. Karpouzis (✉)
Image, Video and Multimedia Systems Lab,
National Technical University of Athens,
Heroon Polytechniou 9, 157 80 Athens, Greece
e-mail: kkarpou@cs.ntua.gr

I. Maglogiannis
Department of Biomedical Informatics, University of Central Greece,
Papasiopoulou 2-4, 35100 Lamia, Greece
e-mail: imaglo@ucg.gr

Keywords Multimedia content adaptation · Personalization · Metadata · MPEG-4 · MPEG-21 · User modelling

1 Introduction

Viewers of digital media, especially live or taped TV broadcasts and movies, are being confronted and becoming acquainted with a series of technological developments in the realm of consumer electronics and gaming that raise the level of functionality they expect when watching TV [1]. Entertainment content such as movies and sports are among the most popular media attractions in the world today, providing a huge potential market impact for related applications. In the forefront of these applications is the provision of enhanced content such as statistics or results from other events in sports broadcasts, and dynamic interactive content such as advertisements or suggestions for related/similar content based on the preferences and usage history of the viewer.

However, most of this functionality cannot be included in the baseline set of services provided by analog TV broadcasts, which are the standard way of consuming TV in most countries, or the MPEG-2 standard used in DVD movies. Besides this, emerging standards which do support these processes such as MPEG-4 and MPEG-7, have yet to enjoy wide commercial adoption, while still not supporting all related features on their own. As a result, one needs to integrate concepts and provisions from multiple, existing standards (so as to not plague the already established content production chain) and deploy them on a PC-like set-top-box for decoding.

The need for this integration led to the development of the MELISA (Multi-Platform e-Publishing for Leisure and Interactive Sports Advertising) framework [13], which involves two types of user terminals, set-top-boxes, and personal digital assistants (PDAs), respectively, functioning with two diverse transmission channels, i.e., Digital Video Broadcasting (DVB) and UMTS, one of the most-publicized 3G mobile phone technologies. Terminals specifications and network characteristics result in inherent diversity with respect to receiving, processing, and displaying multimedia content processes. As a result, in order to comply with the Create Once, Publish Everywhere (COPE) principle, which is considered a prerequisite in modern-day production environments, one must cater for the adaptability of the presentation and the content itself. Besides this, since the integrated system handles, encodes, and presents multimedia content coming from different vendors, the respective intellectual property rights (IPR) must be retained throughout the complete process. These two additional requirements can be dealt with successfully via the inclusion of concepts presented within the MPEG21 framework, introducing additional middleware components which mediate the production, transmission, and playback phases, thereby respecting the established production process; a more detailed description of the initial processlevel architecture is provided in [24].

One of the major advantages of this approach, elaborated in Section 3, is that all incoming content is associated with metadata not just related to playback requirements or parameters, but also with respect to its actual content and available enhancements. As a result, filtering on the receiver side caters for the additional functionality described above by informing the viewer for related content [13]. However, the above-described architecture failed to address one particular requirement: that most people tend to watch TV as a group, thereby turning the whole process into a social activity [18] and making content recommendations more complicated. This lack of support has led a number of researchers to propose different solutions to this, starting from collaborative Web browsing [15] and recommendation of Web pages [2] to suggestion of movies based on their rating from individual users [23]. Zhiwen et al. [29] and Bonnefoy et al. [23] address the problem by collapsing all properties of a user profile and using them to form points in hyper-space which they cluster; this is actually a general purpose approach which may or may not be useful, but in any case fails to model the *dynamics* of a group of viewers by not assigning roles to every one of them. Consider the following example: it is Sunday evening and the match of the day is on TV; a typical family with two teen age children gathers around the television, with one child being a huge fan of football, while the other is indifferent, but will join in just for the fun of it. Besides the football match, a family movie is also available: if one just considers the casual preferences as votes, the family will watch the football match even if the mother would like to watch the family movie and feels very strongly about it. Taking into account different social theories, which Masthoff [18] summarizes in an excellent manner, one may favor the choice of a particular user over that of the majority, in special circumstances (on their birthday, for instance) or to maximize the total, aggregated feeling of pleasure.

This article concentrates on different strategies providing different functionalities when attempting to aggregate heterogeneous content from different providers and adapt its presentation to the preferences of multiple users, utilizing and building on the concepts of the original MELISA system. Here, content-related metadata and user metadata are combined to mimic the intrinsic processes of a group of people and the results of this process when watching a selection of TV or movie programs is discussed. Section 2 describes in a high-level fashion the general architecture of the server- and client-side of the MELISA system, indicating core processes and the flow of content and information. Section 3 explains the concept and properties of a Digital Item, as standardized by the MPEG-21 standard and elaborates on its use in this context, while Section 4 utilizes all the above-mentioned concepts to build and integrate an aggregation strategy for the needs and preferences of all viewers in the group. Section 5 concludes the paper.

2 Architecture of the MELISA integrated system

The system architecture consists of two separate systems: content preparation and management, assembly and packaging of metadata and preparation of the content for transmission is handled at the *sender side*, while the *receiver side* receives the encoded content, decodes and filters it according to any active user profiles, renders any available and/or suitable visual enhancements (e.g., off-side lines in a football broadcast), and handles any interactive components. The content preparation process includes preparing advertising content and designing 2D graphics templates for enhanced content (e.g., statistics) at transmission time. Content adaptation to the user terminal is performed at the server side so that the content is created once, but adapted according to the two targeted networks and terminal types prior to transmission to allow for efficient display and manipulation on the client side.

The overall architecture of the system at the sender side is given in Fig. 2. There are three major sub-systems grouped by numbers, namely Editing and Workflow (1), Service Management (2), Multi-Platform Publishing (3).

Tasks in the Editing and Workflow sub-system include filming of the environment (e.g., the football pitch), off-line scheduling of the event (e.g., scheduling of successive races in an athletics event), preparation of enhanced content templates, and set-up of a repository structure corresponding to the schedule of the events with some basic metadata (name of athlete, nationality etc.). Apart from these offline processing tasks, this sub-module also involves online authoring at transmission time, post-production and workflow control tasks. In the same manner, the Service Management sub-system takes care of any the betting and advertising media, catering for inplay betting, whenever available.

Figure 3 describes the Multi-Platform Publishing which receives all online and off-line content prepared by the other sub-systems and handles the encoding and adaptation processes. In the case of DVB-S broadcast, video is streamed in MPEG-4 over an MPEG-2 Transport Stream (TS) [6,24]. The original video resolution is reduced to fit the lower transmission and playback capabilities of mobile terminals and the related UMTS protocol. Since the targeted receiver architectures offer different degrees of media delivery, interactivity and responsiveness, it is essential to break down both the captured and synthesized material to match the relevant device. Thus, the publishing sub-system prepares different versions of the content for delivery.

3 Digital Items in the proposed framework and implementation of the MPEG-21 standard

In the MPEG-21 [19] framework, applications producing and handling complex digital objects can utilize the Digital Item Declaration Language (DIDL) to declare them. DIDL defines the relevant data model via a set of abstract concepts, which form the basis for an XML schema that provides broad flexibility and extensibility for the actual representation of compliant data streams [14,16].

Using DIDL to this aim caters for creating pre-defined templates during the planning process, which form the initial MPEG4 scene [8] and are instantiated and adapted in real time at broadcast time. In addition to that, the fact that every DI may contain a specific version of the content for each supported platform, reduces the need for assembling and adaptation to be performed only once (during its creation) and not on a per-usage basis.

3.1 Digital Rights Management (DRM)–Rights Expression Language (REL)

The shortage of instruments to express complex usage permissions in an unambiguous, machine-readable way is one of the limitations of DRM technologies and perhaps what hampers wide commercial adoption. In the framework of fusing multimedia content from different providers and distributors, usually with different permissions and usage restrictions, this deficiency hampers establishing business relationships since ensuring proper use is not always straightforward. To tackle this, a process to express usage rights and permissions in machine-readable licenses, guaranteed to be unambiguous and secure, is of utmost importance. The REL [19] is a viable solution to technical interoperability between proprietary DRM systems.

The basic component of a REL is the *rights expression*, which describes all permissions granted to a user of the related protected content, while retaining its integrity via builtin measures to verify authenticity and resist tampering (e.g., unique digital signatures). Via REL, the content author or distributor specifies the parties allowed to use a digital resource (content, service, or software application), the rights available to these users and the terms, conditions, or restrictions necessary to exercise those rights on the particular resource. In this framework, four critical elements are defined: *principals*, identifying entities (persons, organizations, or devices) to which rights are granted, *rights*, which specify activities or actions that a principal can be granted to exercise against some resource (e.g., playback or print), *resource*, defining objects for which principals can be granted a right, and *conditions* that must be met before the right can be exercised. Resources can be in the form of digital works, services or data owned by a principal and identified by a Uniform Resource Identifier (URI). The relation of these elements is such that the proper authority should be able to define *rights* granted by a *principal* for specific *resources* and the *conditions* under which those rights apply. The only concept relevant to content encryption is the provision of processes to ensure that rights expressions are tamper proof and capable of authentication. In this framework, the basic MPEG-21 REL element is the *license*. A license can contain one or more grants [27]; the license issuer that gives the grants that the license contains, and additional administrative information. Each grant must contain information to identify the four elements (principal, resource, right, and condition) associated with it and must be digitally signed by the license issuer.

3.2 Content authoring and management

The system includes a range of authoring tools for production, encoding, and playback of interactive multimedia content in MPEG-4 for a variety of devices over fixed, wireless, and digital television networks [17]. The Multimedia Production Tools incorporate MPEG-4 and MPEG-7 content creation modules for encoding and transmission over DVB.

The platform foresees the infrastructure to support intelligent real time game statistics and enhancements, utilizing information from various sources, both historical and during the events. This approach aims at providing the viewer with valuable information presented in natural way, anywhere, thus increasing their intent in sports broadcasts. Advertising authoring tools aid the production and placement of dynamic advertisement of sports-related and other products. The system allows dynamic scene generation based on predefined templates. The use of these templates allows broadcasters to prepare their visually enhanced and interactive broadcasts well in advance, thus providing this service even during live events. Figure 4 shows an overview of the server-side architecture, adapted to include the necessary concepts for DRM. DIs serve as the system input and are stored in the main repository, including links to binary (e.g., video, graphics, etc.) files. During this authoring process, each contributing entity is recognized for the specific resources that it supplies (MPEG2 video broadcasts, advertisements, visual enhancements, tracking information and betting options). Information on the rights of the different users, based on the available subscription levels are encoded in the form of usage conditions and included in the resulting BIFS [4] file, provided as the output of the multiplexing procedure and transmitted to the end-user terminals.

3.3 Video content authoring

The DI concept as proposed by MPEG-21 is utilized for the downloaded clips or images that are transmitted from the Server to the set-top-box or a mobile device. The creation and usage of the DI can be divided into two parts. The first part is the DI creation that requires the collection of information from automated and user input and the second part is the DI management and streaming to the client platforms (Fig. 1). This methodology is common whether we are dealing with adverting clips or enhanced content replay clips. In the first case the DIs are generated in a pre-broadcasting phase. The advertising unit receives the video clips and the DI items are generated and stored in a central Information Repository [13]. In the latter case the same method is followed, only that this time the process is performed during the broadcasting phase.

3.4 Content filtering for personalization

The system provides end users with the possibility to see only information that they are interested in. One flexible way to perform content personalization, from the technical point of view, is to filter any incoming BIFS Updates; since the same BIFS Updates are broadcast to all clients, filtering is only possible at the client side, i.e., on the set-top-box before display.

MPEG-21 concepts are put to use for personalization and content filtering. The set-top-box of a known (registered) user contains an MPEG-21 DIA (Digital Item Adaptation [21]) Description, which contains the registered user's preferences; when the user terminal receives a BIFS Update, this is filtered according to its genre and the user preferences therein. The main issue here is to find a way to transport synchronously the BIFS Update and its associated metadata indicating its genre, in order to make sure that the Update is not received before its description. One way to achieve this is by grouping the BIFS Update and its genre within a DID, and to stream the complete DID to the clients. This indicates that the Update belongs to the genre "Statistics". Obviously, according to the user preferences in the DIA Description which exists in the system, in this case the Update will be filtered out by the client terminal and therefore not displayed.

The *preferences* elements defined for the MELISA purposes are shown in Fig. 5. These involve preference registration on the sport event or location, favorite broadcasting channel (creation preferences), favorite sports and preference on viewing related sport description, , the types of visual enhancements the viewer prefers to view, the types of bets (if any) the viewer likes to place, the viewer's favorite athlete/teams/statistics, etc., or the language the viewer likes all information to be encoded into. The preference value attribute that accompanies every different preference definition denotes the significance of that definition in the final calculation again of the filtering decision metric.

3.5 Digital Item generation

Two basic APIs were implemented for the creation and management of DIs, the MELISA Digital Item Creator (MDIC) API and the MELISA Digital Item Manager API (MDIM) (Fig. 6). The MDIC utilizes information both from the content author, as well as from automated processes at the sender side. Initially, the content author includes all the necessary information that the DI will contain, i.e., the video content. In this stage the DRM information could be added in the DI, as specified using REL. The system was designed to generate metadata information to describe the content based on MPEG-7. The information describing the video clip is collected from the Information Repository, encoded in XML form, and is included in the DI. The DI with the collected information and the multimedia content (video clip, image etc.) encoded in base64 is then passed to the MDIM; this kind of encoding was chosen in order to ensure compatibility with the MPEG-2 standard and enable streaming over satellite networks.

The MDIM separates the encoded visual information from the XML representation and maintains only a URL reference to it; this is performed to minimize data redundancy and speed up system response. The DI information is also stored in the Information Repository for later use. Should the broadcast director decide to transmit a DI, this is retrieved and then serialized in the form of a Java object. This serialized Java Object is then transformed into a binary array with a description header, ready for FlexMux [20] encoding and transmission.

The encoded and transmitted information is received by the client platform, where it is converted back to its original

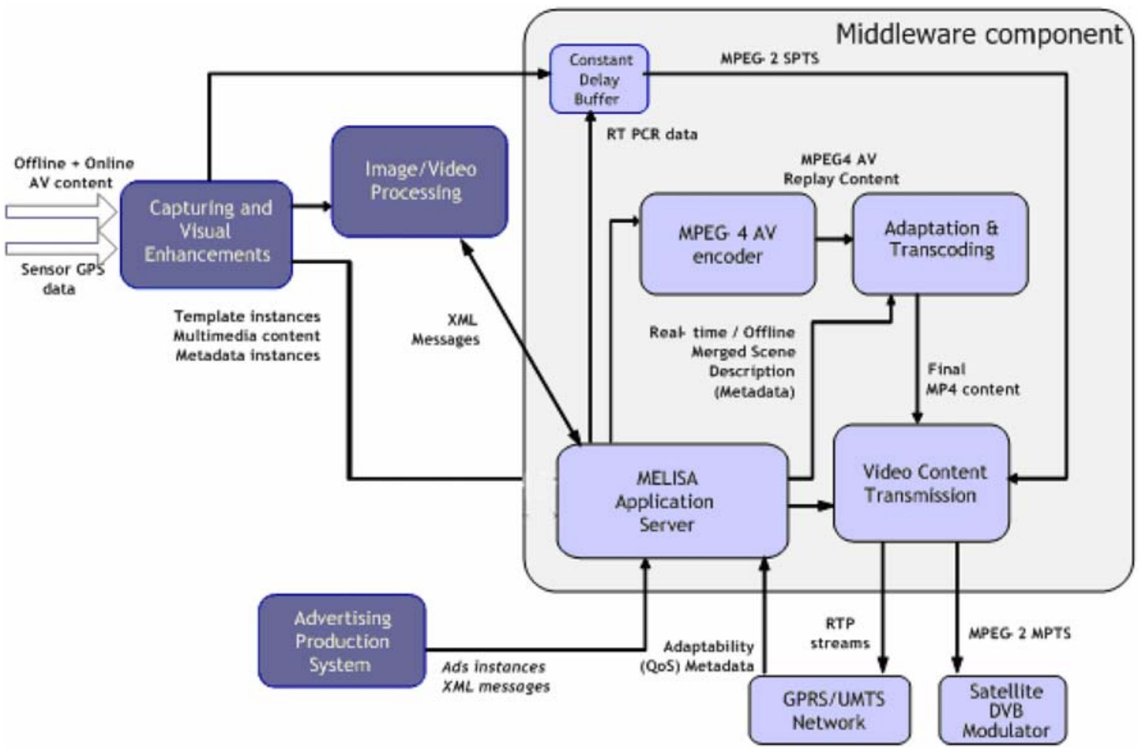

Fig. 1 Overview of the established content production and delivery architecture of the MELISA system [13]

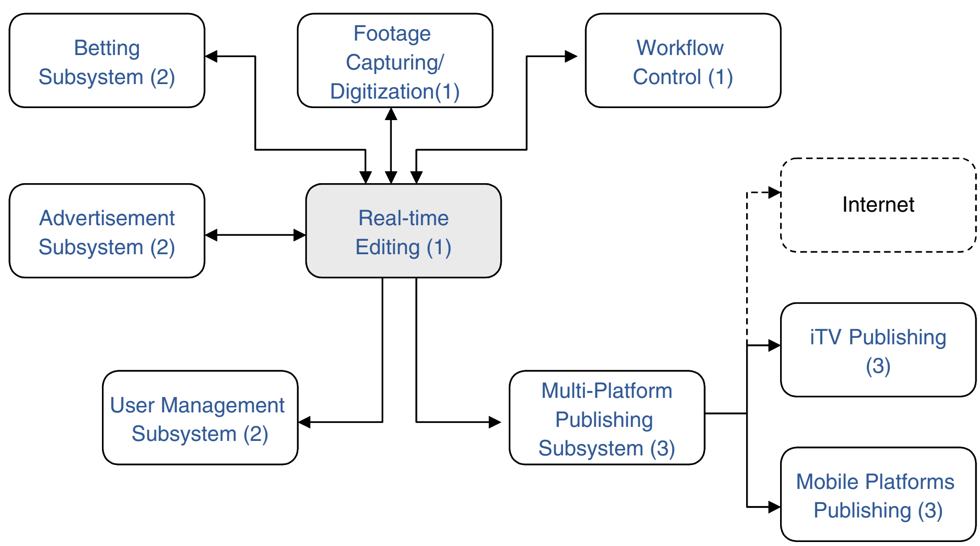

Fig. 2 A process-level schema of the sender architecture [13]

form and then received by the local DIM. User adaptation or Profiling is performed at this stage since all the necessary filtering information is included in the MPEG-7 metadata. The DIM filters the resource and decides whether to reject or accept it. The obvious advantage of having all the metadata in the DI and not in a separate stream is that at the Client side timing issues between the download clips and metadata do no longer exist.

The MDIC API is based on the actual Digital Item Declaration Language (DIDL) XML elements. Every single MDIC class has a corresponding XML element in the schema, as generated using the Castor Java XML Data Binder. We chose the data-binding method so that we can easily read and write valid XML documents. This adds flexibility to our system since any changes to the DIDL schema can be easily incorporated in our system, by simply re-creating the Java classes. The initial system was designed to transmit metadata in the form of serialized Java Objects, so this approach is maintained in the new design. The necessity of working with and transmitting Java Objects was dictated by the set-top-box

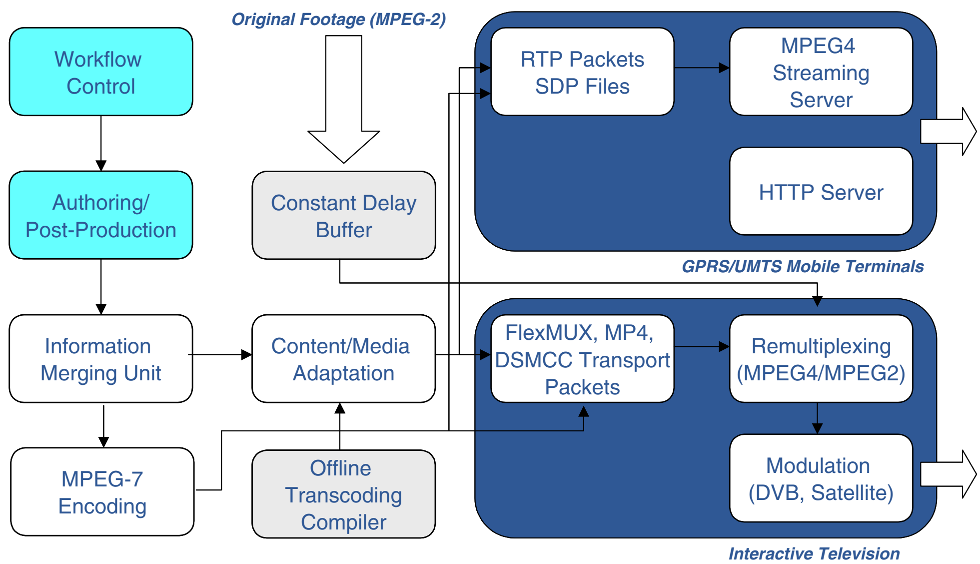

Fig. 3 An abstract schema of the Multi-Platform publishing sub-system (sub-system 3)

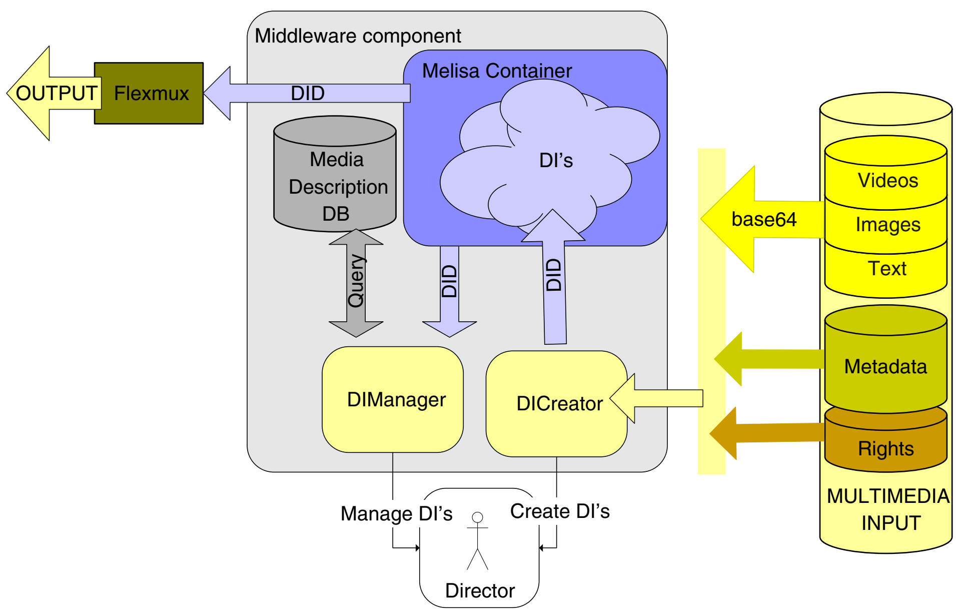

Fig. 4 The server-side architecture with an initial distribution of roles in the Digital Item model

limited processing power. Rather than transmitting XML data and having to parse them to extract the information, the information is already in the Java Objects.

The MDIC architecture can be conceptually divided into three layers (Fig. 7). The first layer contains all the classes that where generated from the XML DIDL schema, while the second one, the *Creator* class, provides methods for the creation of any element of the DIDL schema. This class is responsible for updating the third level element object with the required information and returning the object to the next level layer, the MDIC API, which is responsible for the creation of the DIs. This layer requests the element objects from the Creator class in order to form the Digital Items. The generated Digital Item are then either stored in the Information Repository, or exported to XML format files. Additionally a Digital Item can be provided to the system in XML form, to be imported to the system [13].

The MDIC is integrated in the whole information flow as an agent responsible for handling the offline and online processing of the Digital Items and is connected to the work flow control in order to be able to accept notifications concerning the available content. The Work Flow Control (Fig. 8) receives a notification from the Visual enhancements Unit every time new video content is generated. The Workflow control retrieves the Video Clip and notifies the MDIC (Fig. 7) that a new clip is available. The MDIC operator retrieves the basic metadata information in order to form the Digital Item, i.e., event information, description DI IDs. Metadata infor-

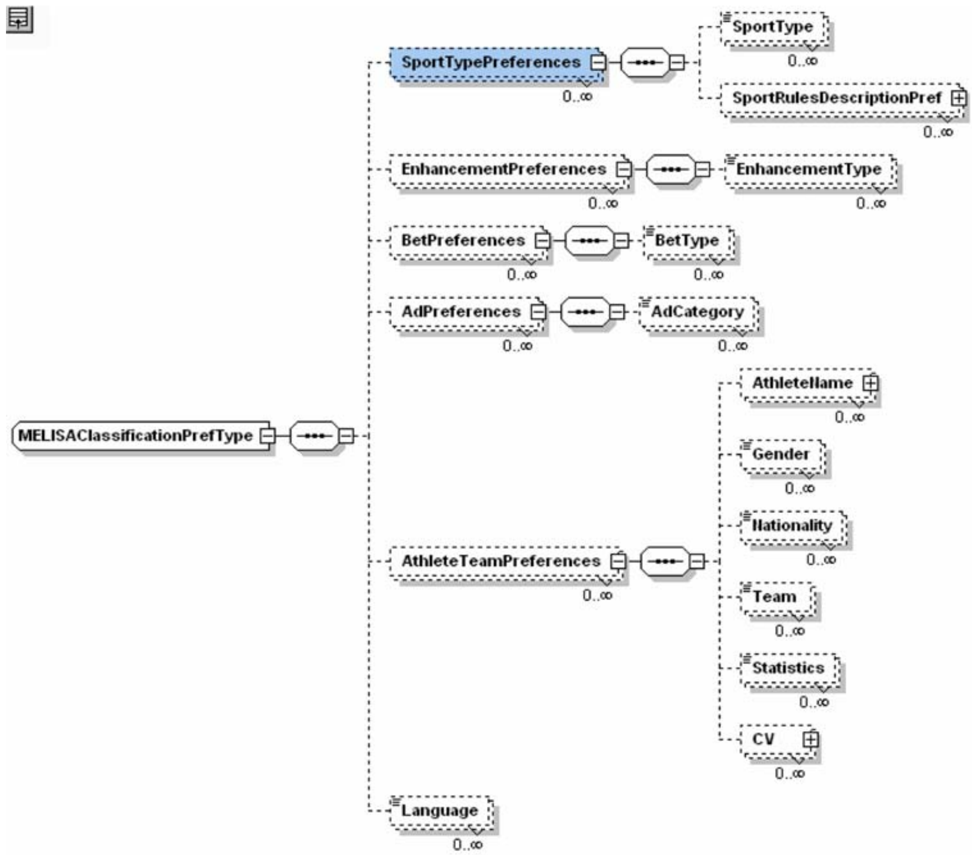

Fig. 5 Preferences elements [13]

mation is requested from the Metadata Manager and optionally from the MELISA Digital Rights Manager (MDRM). The information is then passed to the MDIM where it is stored in the Information Repository. The MDIM notifies the Work Flow Control that a new DI is ready. The director can see the list of the available DIs, and decide whether he will broadcast it.

3.6 Client-side application

The receiver platforms supported by the system are high-end set-top-boxes, portable digital assistants and Java MIDP-enable mobile phones. The system provides content adaptation according to terminal capabilities, e.g., adapted interaction, visual presentation in high or reduced resolution graphics etc. During this process, the receiver initially decodes the incoming encoded streams to gain access to the stored visual information and the associated metadata. As stated earlier, the Digital Items consist of visual information in the forms of MPEG-2 video and MPEG-4 graphics, multiplexed with MPEG-7 metadata and MPEG-21 information into an MPEG-2 stream to be transmitted.

Each broadcast event can be conceptually divided into segments, according to the program planning performed by the broadcaster. Every event is planned in advance and divided into segments, e.g., the first half of a football game, the 15 minute advertising break during half time, the first section of a Grand Prix, the 5 minute advertising breaks during the race etc. Each segment can be described using a Digital Item that basically contains the MPEG-7 metadata description.

The stream is received by the set-top-box and the FlexMux stream [24] is decoded by the FlexDemux module. The DI and the incorporated MPEG-7 metadata contain the information necessary to describe the resource and the related rights. Since this is a multicast environment, rights enforcement and profiling has to take place in the receiver, after the decoding process. Essentially, if the user (principal) logged on to a particular terminal has the rights necessary to perform specific actions on the included content, and his personalized profile allows it, the set-top-box container will forward the binary and text data to the user interface for playback. In the case of offline video replays (downloaded clips), these are streamed as Digital Items, de-multiplexed when they arrive at the set-top-box; the local MDIC then takes over to handle the Digital Items. The local Profile Manager is consulted

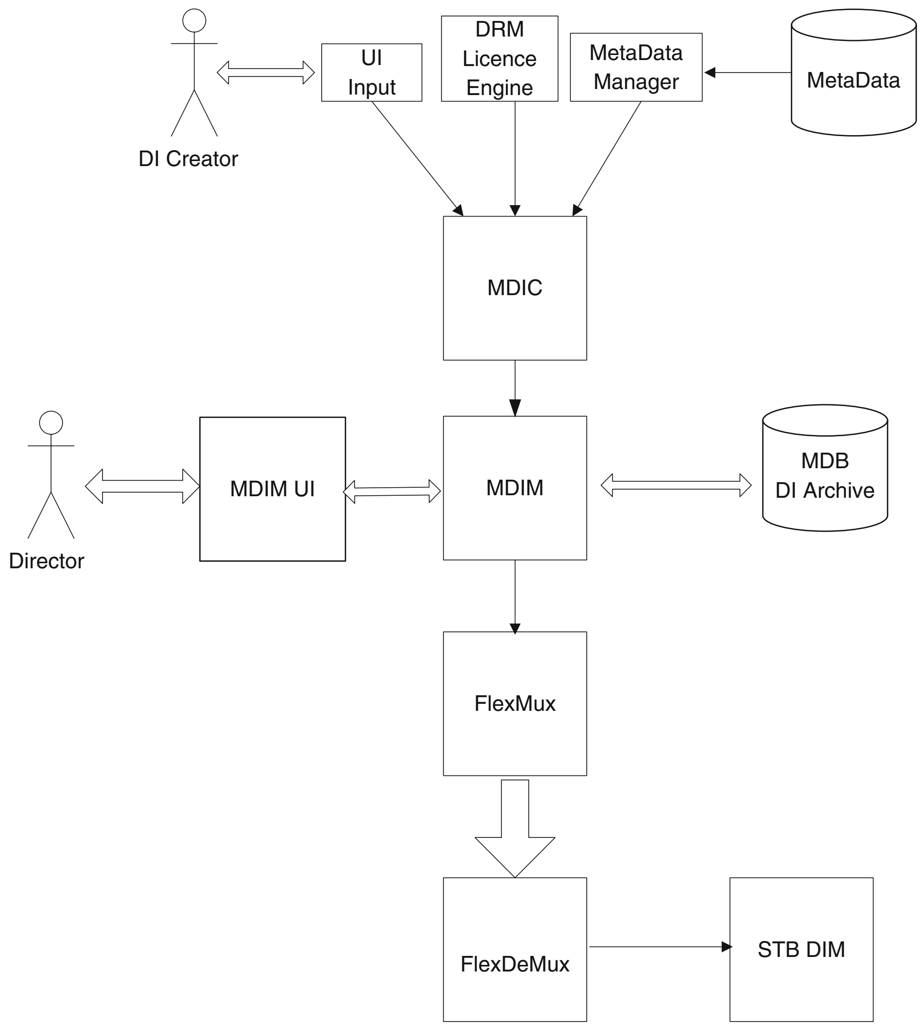

Fig. 6 Logical diagram and data flow during the authoring, encoding and reception processes

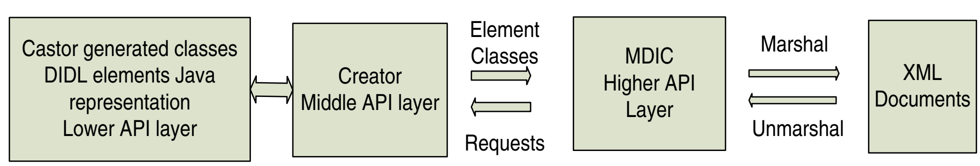

Fig. 7 MDIC architecture diagram

to make a decision on whether the user would like to view this type of content. If the local profile allows the playback of such content, it is displayed to the user; otherwise, it is ignored.

4 Strategies and functionality of multi-user profiling

Masthoff [18] tackles the problem of a group of people watching TV and being offered different alternatives from a social point of view, mentioning the construction of a 'social welfare function' as proposed by social choice or group decision making papers, not necessarily connected to entertainment or even computers. This article mentions a number of examples, from multi-agent systems to database middleware, with one common parameter, '*to take decisions that are not only rational from an individual's point of view, but also from a social point of view*'.

In our work, we implemented several of the proposed approaches from that paper, solving or choosing not to tackle some of the issues mentioned there. For example, in the framework described here, the receiver *knows* who is in the group of viewers at all times, either by explicit input via user cards or by selecting registered users via an on-screen interface. Also, individual preferences are provided by each user in their profile; no attempt to track viewing habits is made. Finally, available recommendations are shown as a sorted list of suitable programs, which can be activated via a button on

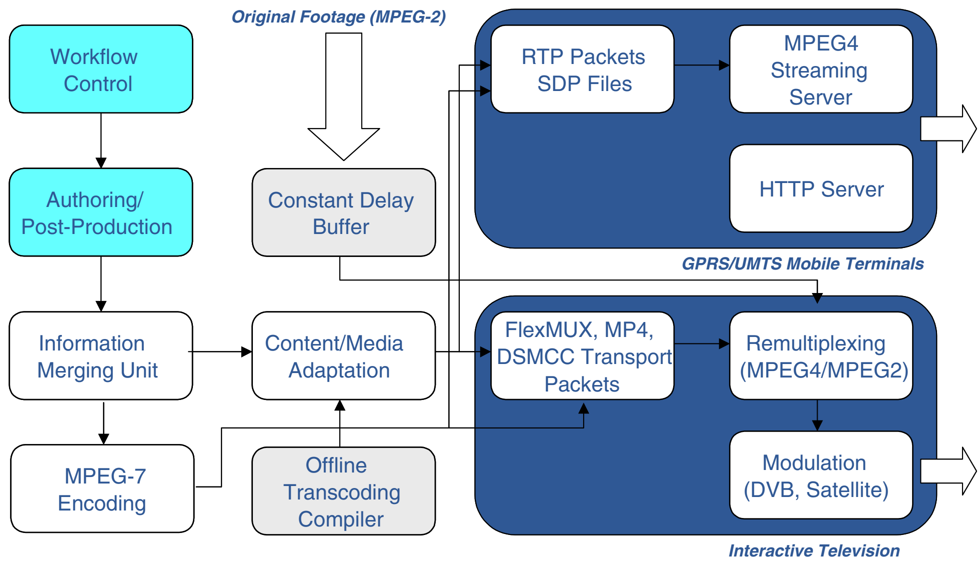

Fig. 8 An abstract schema of the sender sub-system

the remote control. As a general rule, multi-user recommendation systems make a number of assumptions and follow a general baseline:

- If all participants are indifferent with respect to the available choices and just one (say, viewer A) prefers one particular program or method of presentation, then A should be humored (a variation of the Pareto rule [25]).
- When a new user (user A) enters the group and their disposition follows the already established ranking (e.g., Sports over a Movie), then this sequence should be at least preserved. The interesting case has to do with Sports and Movie having matching initial ranks: then, when A joins the group and prefers Sports, the ranking of these two choices may or may not change (to cater for a majority vote, for instance, when A is with the minority), but if it does, it should be Sports > Movies.
- The Condorcet criterion [7] mentions that '*An alternative x is a Condorcet winner if for each other alternative y: x is preferred to y by the majority of individuals*' [18]. However, in the case of TV viewing, this criterion may not be suitable, since participants in the minority may object strongly to alternative x, thereby reducing the aggregate feeling of content in the process. As a result, Masthoff proposes that in the selected framework '*each individual's satisfaction with the results should be above a certain threshold*'.

Different aggregation strategies are available, with respect to the above assumptions. In *plurality voting*, which is very often the first way to go, votes for the different alternatives are tallied and the programs are sorted with respect to that result. When the available alternatives vastly outnumber the viewers, it is possible (and usual) that the voting process will result in a tie; this situation (which, in practice, may present harder problems to solve) may be alleviated by allocating multiple (e.g., three) votes per person, which usually results in an alternative being voted by multiple viewers as a 'second choice'. This approach is usually based on a 'binary' vote: participants either vote for an alternative or they do not.

If viewers are presented with the possibility to *mark* the available alternatives with marks, say, from 1 to 5, or 1 to 10, then these ratings can be added or even multiplied, and the larger the sum the earlier the alternative that will appear in the sequence. Since these marks effectively correspond to the expected 'pleasure' that the particular alternative will bring to the viewer, these strategies are called *utilitarian*. Especially in the case of large groups, people in the minority always miss out; in smaller groups and in the case of adding the votes, each individual has a larger share in the final decision. A variant of the marking system, combined with ranking, is the *Borda count*: here, the viewer ranks the alternatives and the one at the top receives most points (say, five for a list of five programs), while the one at the bottom receives zero points. A final strategy, which is also quite popular in practice, offers the opportunity to the viewers to vote for as many alternatives they like, effectively indicating their preference to one or more genres or presentation options; choice is determined by the number of votes, while a 'cut-off' criterion of a threshold of votes hampers the display of the least popular options.

More or less, the above-mentioned options do not take into account the *social* processes or any particular requirements or occasions (cf. the birthday example mentioned earlier). One could subscribe to the idea that 'a group is as happy as its least happy member', so ranking should take into account the minimum of the individual ratings. In essence, the *least* *misery* approach results in all members of the group enjoying one minimum of satisfaction; if user A has seen a particular movie and does not want to see it again, the movie will not be shown, even if all the others would like to see it; an alternative of that is to take out alternatives which score below a particular threshold for each user. In the middle between the 'least misery' strategy and the 'most pleasure' strategy which takes into account the maximum of each participant's rating, lies the *fairness* strategy: here, best-marked items from all viewers are selected, so even if one has to watch something they do not like, this feeling is reduced by knowing that in the following, something they like will be broadcast.

Moving away from the above approaches, which more or less attempt to take into account what *every* user would like to watch, the *Most Respected Person Strategy* offers the remote control to the 'leader' of the group and effectively the choice of genre and presentation options. Special occasions, like the 'birthday' example fit in this area; in the case of a family viewing a selection of programs, usually the choice is made by one of the oldest participants or the one that feels more strongly about an option. One may also assign weights to the preferences of the participants, either by taking age into account, or the fact that they were humored during an earlier session, or even the degree of savvy in a particular context (e.g., familiarity with football during a World Cup broadcast).

5 Conclusion

The proposed system was originally designed to handle text, image, and live video content in a conventional way; injection of MPEG-21 concepts showed that the adoption of such a standard brings a number of advantages, both related to technical issues such as streamlining the production and delivery line, as well as preserving the intellectual property of the different contributors and allowing for proper usage. The integrated rationale of content delivery and presentation offers a unique, personalized performance to a single viewer; however, to be able to adapt and adjust to a group of people, with different disposition towards different genres and means of presentation, one must utilize concepts from social theory and model the particular group dynamics.

This article builds on the technical concepts presented in [13] and exploits aggregation strategies [18] to filter the available content and presentation options with respect to what the group as a whole would like to watch. This is not a straightforward task, nor one solution is possible or best-performing; instead, different approaches are discussed, each complying with a different strategy and performing better with a particular metric in mind (e.g., to provide the least misery for all viewers in the group). The choice of solution to be followed should in conclusion depend on the particular group or left to the viewers to decide.

Acknowledgments This work was initiated within the framework of MELISA (http://melisa.intranet.gr), an EU-funded fifth framework IST project (IST-2001-34755); the multi-user and aggregation functionality was researched and developed within the framework of the FP6 IP CALLAS, Project IST-34800 (http://www.callas-newmedia.eu/).

References

1. Abdeljaoued, P.Y., Christopoulos, C., Ebrahimi, T.: Universal multimedia access from wired and wireless systems. Birkhauser Boston Trans. Circuits Syst. Signal Process., Spec. Issue Multimed. Commun. **10**(3), 387–402 (2001)
2. Bollen, J.: Group user models for personalized hyperlink recommendations. In: Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, LNCS 1892, pp. 38–50 (2000)
3. Bonnefoy, M., Bouzid, N., Lhuillier, K.M.: "More Like This" or "Not for Me": Delivering Personalised Recommendations in Multi-user Environments, LNCS 4511, pp. 87–96 (2007)
4. Concolato, Dufourd, J.-C.: Comparaison of MPEG-4 BIFS and some other multimedia description languages. In: Proceedings of WEMP Workshop (2002)
5. ConcolatoDufourd, J.-C., Moissinac, J.-C.: Creating and encoding of cartoons using MPEG-4 BIFS: methods and results. IEEE Trans. Circuits Syst. Video Technol. **13**(11), 1129–1135 (2003)
6. Coding of Audio-Visual Objects (MPEG-4)—Part 1: Systems, ISO/IEC 14496-1:2001
7. Condorcet, M.de.: Essai sur l'application de l'analyse a' la probalite des decisions rendues a' la pluralite des voix, Paris (1785)
8. Dufourd, J.-C.: BIFS: scene description. In: Pereira, F., Ebrahimi, T. (eds.) The MPEG-4 Book, Chap. 4. IMSC Press, New Jersey (2002)
9. Holtman, K., Mutz, A.: Transparent Content Negotiation in HTTP, IETF RFC 2295 (1998)
10. ISO/IEC JTC1/SC29/WG11/N3382 14496-1:2001 PDAM2 (MPEG-4 Systems), Singapore, March 2001
11. Jameson: More than the sum of its members: challenges for group recommender systems. In: Proceedings of the International Working Conference on Advanced Visual Interfaces, pp. 48–54 (2004)
12. Ji, M., Shen, S.M., Zeng, W., Senoh, T., Ueno, T., Aoki, T., Hiroshi, Y., Kogure, T.: MPEG-4 IPMP extension for interoperable protection of multimedia content. EURASIP J. Appl. Signal Process. **14**, 2201–2213 (2004)
13. Karpouzis, K., Maglogiannis, I., Papaioannou, E., Vergados, D., Rouskas, A.: MPEG-21 Digital Items to support integration of heterogeneous multimedia content. Comput. Commun. **30**(3), 592–607 (2007)
14. Kofler, I., Seidl, J., Timmerer, C., Hellwagner, H., Djama, I., Ahmed, T.: Using MPEG-21 for cross-layer multimedia content adaptation. Signal Image Video Process. **2**(4), 355–370 (2008)
15. Lieberman, H., Dyke, N., Vivacqua, A.: Let's browse: a collaborative Web browsing agent. International Conference on Intelligent User Interfaces (IUI99), pp. 65–68, ACM Press, New York (1999)
16. Lopez, F., Martinez, J.: Multimedia content adaptation modelled as a constraints matching problem with optimisation. In: Proceedings of WIAMIS '07. Eighth International Workshop on Image Analysis for Multimedia Interactive Services, pp. 82–88 (2007)
17. Maad, S.: Universal access to multimodal ITV content: challenges and prospects, User Interfaces for All, pp. 195–208 (2002)

18. Masthoff J.: Group Modelling: Selecting a Sequence of Television Items to Suit a Group of Viewers, User Modeling and User-Adapted Interaction, vol. 14, pp. 37–85. Springer, Berlin (2004)
19. MPEG-21 Overview v.5, ISO/IEC JTC1/SC29/WG11/N5231, Shanghai, October 2002. http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm
20. MPEG-4 Overview v.21, ISO/IEC JTC1/SC29/WG11 N4668, March 2002. http://www.chiariglione.org/mpeg/standards/MPEG-4/MPEG-4.htm
21. Multimedia Description Schemes Group, MPEG-21 Digital Item Adaptation: Providing the tools for managing audiovisual content and services in today's diversity of networks, devices and user preferences, White paper on MPEG-21 Digital Item Adaptation. http://www.chiariglione.org/mpeg/technologies/mp21-dia/index.htm, last retrieved: September 10, 2008
22. Munchurl, K., Jeongyeon, L., Kyeongok, K., Jinwoong, K.: Agent-based intelligent multimedia broadcasting within MPEG-21 multimedia framework. ETRI J. **26**(2), 136–148 (2004)
23. O'Connor, M., Cosley, D., Konstan, J., Riedl, J.: PolyLens: a recommender system for groups of users. European Conference on Computer-Supported Cooperative Work, pp. 199–218. Bonn, Germany (2001)
24. Papaioannou, Karpouzis, K., de Cuetos, P., Karagianis, V., Guillemot, H., Demiris, A., Ioannidis, N.: MELISA—a distributed multimedia system for multi-platform interactive sports content broadcasting. In: Proceedings of EUROMICRO Conference, pp. 222–229 (2004)
25. Pareto, V.: Cours d'economie politique. Lausanne, Rouge (1897)
26. Seshan, S., Stemm, M., Katz, R.H.: Benefits of transparent content negotiation in HTTP, Global Internet Mini Conference, Globecom 98, November 1998
27. Wang, X.: MPEG-21 Rights Expression Language: enabling interoperable digital rights management. IEEE Multimed. **11**(4), 84–87 (2004)
28. W3C XML Path Language (XPath) Recommendation, November 1999. http://www.w3.org/TR/path
29. Zhiwen, Y., Xingshe, Z., Yanbin, H., Jianhua, G.: TV Program Recommendation for Multiple Viewers Based on user Profile Merging, User Modeling and User-Adapted Interaction, vol. 16. Springer, Berlin, pp. 63–82 (2006)