Un Eco-système de Médias Réseaux basé sur la Conscience du Contexte, du Réseau et du Contenu à destination de l'Internet Media du Futur

Daniel Négru

Université de Bordeaux 1
Laboratoire Bordelais de Recherche en Informatique

HABILITATION A DIRIGER DES RECHERCHES

Au titre de l'école doctorale de Mathématiques et Informatique de Bordeaux

Présentée par
Daniel NEGRU

Titre :

A Networked Media Ecosystem based on *-Awareness for Future Media Internet

Un Eco-système de Médias Réseaux basé sur la Conscience du Contexte, du Réseau et du Contenu à destination de l'Internet Media du Futur

A soutenir le 05 Décembre 2013 devant la Commission d'Examen composée de :

Pr.	**Ahmed**	**KARMOUCH**	**Rapporteurs**
Pr.	**Pascal**	**LORENZ**	
Pr.	**André-Luc**	**BEYLOT**	
Pr.	**Gerardo**	**RUBINO**	**Examinateurs**
Pr.	**Djamal**	**ZEGHLACHE**	
Pr.	**Francine**	**KRIEF**	

Summary

Today's trend and absolut necessity is to build the Future Internet. Many research actions are headed towards this and most of them outline the Media domain as one of the most predominant and essential to be considered. Indeed, Media content is already dominating the Internet traffic and is continuously increasing whereas the current architecture had not been built with Media in mind. Efficient delivery of Media content becomes now a key evolution not only for pleasing End-Users but also (and consequently) for revealing new business opportunities to all the actors in the value chain.

In this thesis, we propose a new architecture for Future Media Internet, centred around the concept of a win-win collaborative approach between all the stakeholders. The "Networked Media Ecosystem" we aim to construct will enable new features and benefits for all stakeholders, essentially thanks to the innovative *-Awareness features it enables and their outcomes in terms of quality enhancement and adaptation. The proposed approach relies on: (1) a virtual Home-Box layer, which empowers Context- and Network-Awareness at the application level, linked with an efficient content delivery system, and provides related information for permitting to achieve Content-Awareness at the network level; (2) a virtual Content-Aware Network (CAN) layer, which role is to consider the information related to the transported content and take proper actions related to it, in addition to providing network-related information to the upper layers. The proposed architecture has not only been designed and simulated but also conceived, implemented and validated through the European Project named ALICANTE. It provides interesting results and has paved the way towards future evolutions and research directions in the field of Networked Media.

Table of Contents

Table of Figures

Table of Tables

Chapter 1

Introduction

Thanks to its rapid advance and exponential growth during the last years, the Internet is now playing a key role in everyday life, in information exchange, business, administration, education, social life and entertainment. As well, audiovisual services are used in more and more domains; Media Services are now dominating the global Internet traffic – at a percentage predicted to reach 90% by 2014. Citizens are increasingly using the Internet for accessing and sharing multimedia content, a trend leveraged by the increase in access bandwidth and advances in service technologies.

However, the current Internet architecture and technologies present significant weaknesses with regard to provision of Media Services. The Internet, initially designed for resilient, opportunistic transport of non-real-time data, does not provide adequate mechanisms for the differentiation, appropriate handling (i.e., content-aware processing, QoS, routing/forwarding, security, monitoring) of multimedia traffic. As a consequence, business opportunities related to Internet-based media provision are seriously hampered. Quality guaranteed services are rarely offered over the Internet due to the lack of QoS mechanisms, being mostly restricted in "fenced", managed networks with centralized resource management. As a consequence, current Future Internet (FI) debates have shown that a revisit of the architecture and support technologies is needed to allow (among others) efficient distribution of media content.

Another important trend in Internet usage is the high users' expectation for seamless access to services. Indeed, with the compelling proliferation of mobile devices (e.g., smartphones, tablets, laptops, etc.) and the rapid growth in mobile networking technologies (mobile network connection speeds doubled in 2010: globally, the average mobile network downstream speed in 2010 was 215 kilobits per second (kbps), up from 101 kbps in 2009 [1]), users are willing to access high quality services anywhere, anytime and through any device or network. From the End-User's point of view, only the service value counts, not the networked device or software components that implement it.

If these new trends in Internet usage seem to be very attractive, they also impose stringent requirements on the current Internet infrastructure in terms of scalability, reliability and performance. It is especially true while dealing with video services which generate huge traffic that heavily loads on the Internet. Moreover, they are known to be time-critical and loss-sensitive. These services increase the pressure on network infrastructure typically for high capacity, low latency and low-loss in the communication paths.

In such a computing context, it is then obvious that Internet has to growth in its functionality, capability and size to enable the creation and efficient distribution of novel advanced rich Media Services. Indeed, many limitations have been identified in the current Internet [2] such as facilities for large-scale service provisioning, management and deployment, facilities for network, device and service mobility, facilities to seamlessly use context information to enhance and improve existing services and deploy new ones, facilities to support Quality of Service (QoS) up to Quality of Experience (QoE) and Service Level Agreements (SLA), trust management and security, privacy and data-protection mechanisms of distributed data and so on.

The extensive use of the Internet for the distribution of multimedia content, previously provided solely via one-way broadcasting platforms, has led to drastic evolutions in the **Networked Media Content Value Chain**, especially due to the open, interactive and multi-service nature of the Internet. The original networked media content value chain composed essentially of the Content Creator, the Content Provider, the Content Distributor (assuring means for delivery through e.g., a broadcasting network) and the End-Users (Content Consumers) had to be extended for including existing Internet players, such as the Service Providers (assuming, among others, Content Distributor's role) and the Network Providers (or Operators), in order to be able to handle this evolution. One simplified representation of the current media value chain can be seen in **Figure 1**. Of course, today, extensions of such general actors have come out to provide additional or different features (e.g., CDN providers, Cloud providers, etc.).

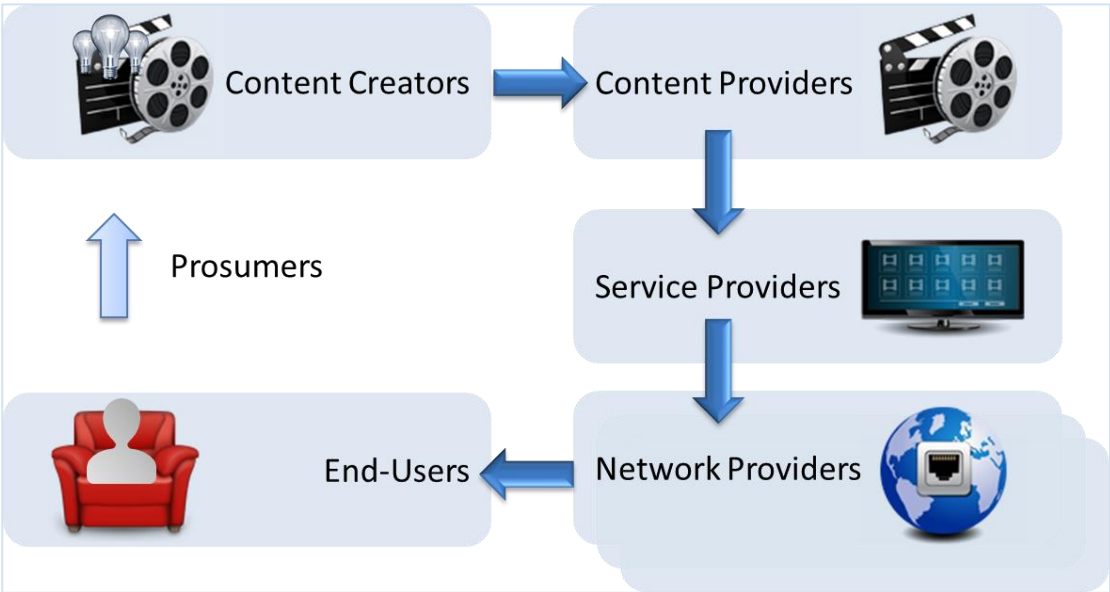


Figure 1: The current Networked Media Content Value Chain (Simplified Representation)

It is also anticipated that the Media Content Value Chain in the Future Internet will involve an additional large number of participants/actors, among which Commercial Providers of contents and services, End-Users acting as both content consumers and creators using various terminals of different capabilities, and Network Operators with virtualisation capabilities. Similar to the notion of "ecosystem" in ecology and business, the concept of "Media Ecosystem" corresponds to a novel collaborative paradigm for sharing and consuming Media Services, dynamically exploiting content and resources from the involved actors. Vital features of such an ecosystem would be seamless integration with the Future Internet architectures, QoS/QoE provision, synergy between the service and multi-domain network planes, maximum flexibility to support a variety of business models, expandability, scalability, and security.

To summarize, the Future Internet is expected to be a communication and delivery ecosystem, where media occupies an important part. The Future Media Internet Architecture – Think Tank group (FMIA-TT) [3] aims to specify a high-level reference model of a "Future Media Internet Architecture" which covers delivery, adaptation/enrichment and consumption of media within the Future Internet ecosystem. The objective is to reach compliancy between architectural works in this domain by relying on some main concepts. Consequently, the FMIA-TT has defined a high-level FMI network architecture based on three layers: *Information Overlay* comprising intelligent nodes and servers with knowledge of the content/web-service location/caching and the network instantiation/conditions (the nodes can vary from P2P peers, secure routers or even Data Centers); *Distributed Content/Service Aware Overlay* containing nodes which filter content and Web services; *Service/Network Provider Infrastructure* acting as the traditional layer of services offered by the ISPs. The users can be providers and/or consumers (prosumers) of the services offered by this layer.

Considering the abovementioned studies, our work introduced a novel architecture aiming to facilitate the deployment of an integrated Media Ecosystem within Future Internet, where all involved actors (Service/Content Creators and Providers, Network Operators/Providers and End-Users) collaborate for the efficient distribution of rich Media Services, both provider- and user-generated ones. The proposed 'Media Ecosystem' architecture therefore includes three environments (the User Environment, the Service Environment and the Network Environment), each one grouping the components of the designated actor, and introduces two novel virtual layers: the Home-Box layer and the CAN (Content-Aware Network) layer. The Home-Box virtual layer is composed of components from the upper layers whereas the CAN layer is composed of components located at the networking layer. The concept around which those two novel virtual layers are being conceived is centred on the efficient cooperation of actors by bringing:

1. Context- and Network-Awareness at the Application level (i.e., enabling Context-/Network-Aware Applications);
2. Content-Awareness at the Networking level (i.e., enabling Content-Aware Networks);
3. Efficient media delivery, including adaptation at both levels towards best possible Quality of Experience.

Considering the state-of-the-art in multimedia oriented platforms, the *-Awareness feature is considered as a principal characteristic of the Media Ecosystem. Currently, there are several terminologies that are used in the literature and in associated works in a non-uniform way and have overlapping semantics to indicate a different "networking" behaviour with regards to the current Internet paradigm: Content-Aware Networking (CAN), Content Oriented Networking (CON), Content-Centric Networking (CCN), Service Oriented Networking (SON) [4][40][6]. More specifically:

1. CAN can be considered as a very general characteristic meaning that the overlay network may have one basic Content-Awareness feature or a more elaborated one, as below:
    a. ***Content-type awareness:*** it performs (in the network elements) appropriate processing for flows having different content-types (e.g., assuring QoS in multiple domain planes, forwarding, adaptation, security, etc.), based on the content-type learned from the packets themselves and/or in combination with

the Management and Control Plane information. For example, a set of VoD flows is considered as having VoD-type and only quantitative differences might exist between individual flow requirements. In our proposed architecture, the content-type awareness concept allows aggregated processing in the Virtual CANs;

b. ***Content-objects awareness:*** named CON in [9], it has a finer granularity; CON assumes that content-objects are individually recognized and processed accordingly (e.g., name-based routing, content-object caching, etc.) [7][8]. It basically deals with *content objects* upon which it applies the following operations: *naming, locating/routing, delivery/dissemination, and in-network caching.* Here, the routing is based not on “location” as in classic IP but on “name”. The CCN approach [4] is a particular case of CON.

The approach (a) has less granularity, but has the advantage of permitting an incremental evolutionary architecture and implies less additional overhead in the routers. A straightforward mapping of VCANs onto MPLS/DiffServ technologies is possible.

The approach (b) has a finer granularity, and it can be seen as a “more revolutionary” one. However, it involves much more processing tasks in the network nodes (and significant additional memory if caching is used). So, media object awareness in the routers requires much more resources and creates scalability problems not yet clarified. Also, the Management and Control Plane is significantly different, (w.r.t. the previous approach) not only at network layer but also at the service layer.

2. SON represents the awareness at network overlay level about the services (service is seen as content plus procedures to manipulate it). In this sense, SON is overlapping with CAN and CON.

In our proposed architectural solution, the basic approach for architectural specification, design and implementation is inspired from option 1, *a* (content-type awareness). The proposed solution consists of the introduction of content-related metadata information for transport, called **Content-Aware Transport Information (CATI) inside media flows** by SP/CP and recognition of these information by the network nodes in order to perform their functions accordingly (based on pre-defined agreements).

Additionally, **user Context-Awareness** is considered (i.e., awareness of the user status/location along with user-terminal-related information). It is a key technology enabling ubiquitous access to services. This feature is all the more important in the case of Media Services since they are very context-sensitive. The promise of Context-Awareness is to provide computing frameworks that track context information from different sources in the network, model it in a way that enables their processing by software entities, understand enough on it and, finally, draw conclusions on it. The latter will trigger **adaptation** decision on the applications behavior. The aim is to provide End-Users with contents, resources and services that suit their needs and preferences without explicit intervention from them. Context could be a user location, preferences and activity, device capabilities, network conditions, environment information such as time, light intensity, motion, sound noise level, etc.

Although different **context-aware systems** have been proposed, the design and deployment of such systems that scale to the size of the Internet and support different Media Services is still a challenging issue. In most current systems, the set of commonly used context information is still limited to identity and location. In addition, these systems are usually tackling specific applications and domains making their extension difficult and almost impossible. An efficient **context-aware framework** requires two main careful designs. First, the context has to be represented in a formal model that will allow it to be, on one hand, processed from software entities and, on the other hand, to be flexible, extensible and interoperable. Second, the context management (context acquisition, processing and dissemination) should also be designed with keeping in mind scalability, extensibility and framework responsiveness issues, all dedicated to always reach the best possible Quality of Experience for the End-User.

We propose a solution that combines Context-Awareness and Content-Awareness in a common framework. The *–Awareness information delivered by User Environment and Service Environment (see chapter 2 for an architectural view) can then be exploited by the Network Environment and associated modules for efficient delivery of adapted Media Services.

Concerning the **delivery** of Media Services, two candidates and competitive used approaches are Content Delivery Networks (CDNs) and P2P systems. In CDNs, the contents are pushed from origin servers to multiple powerful servers – so-called surrogates – with high storage and upload connectivity, deployed at strategic locations of the Internet edges, thus allowing Service Provider to handle much more End-Users requests and avoid typical congestion caused by flash crowds (i.e., a larger than anticipated set of users attempting to access a just published content). However, if CDNs enhance services scalability and network latencies, they also induce heavy scaling cost, especially with the continuously growing Media Services popularity. More, if this approach avoids congestion on the core network, it still has no control on the last mile network (aggregation and access network).

On the other hand, P2P systems can be considered as the logical extreme of the distributed approach for content delivery and, accordingly, are highly scalable. However, if the P2P systems represent a promising low cost approach for highly scalable video content distribution, they also present some weaknesses such as lack of control, high peer churn and unfairness and significant imbalance between the uplink and downlink capacity. These weaknesses may rapidly result on system saturation and poor quality.

A hybrid approach, called also peering CDN or adaptive CDN in some literatures, consists in combining the two approaches to benefit, on one hand, on the reliability of CDN and, on another hand, on self-scalability of P2P networks. This solution brings the content closer to End-Users than CDNs could do and this with a much lower cost. However, the efficiency of this type of solution is highly dependent on how the contents are cached among peers and from where users access the contents they request. Efficient and well designed caching and server selection strategies are then required.

Our proposal for Media Services distribution is in line with the last approach. The idea is to leverage the participating and already deployed Home-Boxes (HB) caching facilities and uploading capabilities to achieve service performance, scalability and reliability, especially in current context where the broadband providers are heavily investing to build out their high speed last mile networks. The proposed architecture keeps the high control on traffic design and management of CDNs in the core network while taking advantage of the user environment capabilities (e.g., in terms of bandwidth, processing, availability, and storage), as P2P solutions do. **This solution brings the contents closer to End-Users** decreasing thus path latency **towards a better Media experience**. An efficient online popularity-based content placement and replacement in the HBs' caches is also proposed. The caching strategy aims to efficiently spread the most popular contents among the HBs' caches in order to make them available to much more users, while keeping lower cost distribution. The caching is combined with an efficient context- and network-aware anycast-based server/peer selection that aims to always provide users with contents from the optimal locations.

In conclusion, we propose a complete and open architecture able to accommodate (1) the current and future needs of media content oriented services and (2) the flexible, scalable and efficient usage of network transport resources over heterogeneous networking technologies. Our architecture instantiates the concept of a "Networked Media Ecosystem" and empowers those features through the creation of two virtual layers, the Home-Box layer enabling context-/network-aware applications and efficient delivery and the CAN layer enabling content-aware networking, both considering adaptation as the key feature for achieving the best possible Media experience.

The contributions presented in this thesis (and elaborated throughout the last five years) consist of:

- The proposal of a Networked Media Ecosystem architecture towards Future Internet [10], with new approaches for actor's cooperation. Details are provided in Chapter 2. An European FP7 IP Project, named ALICANTE [11] has been launched based on this approach and four PhD students have performed their thesis under my supervision based on this concept;
- The proposal of a Context-Aware Framework for Media Services Provisionning, in order to achieve Context-Awareness features. It includes investigations around two models (markup-based and ontology-based) and presentation of the QoE assessment model and tool able to populate the contextual model and to perform evaluation of the quality perceived by the End-User, in order to launch adaptation actions, if necessary. One PhD student and several Internship Engineers have worked on the Context Models also in collaboration with Sogeti R&D department. Concerning the QoE model, one PhD student and a Research Engineer have worked on it, in strong collaborations with two SMEs and EPFL. Details are provided in Chapter 3;
- The elaboration of the virtual Home-Box layer enabling Context- and Network-Awareness at the Application level and the related delivery and adaptation features towards maximisation of the End-Users' Media experience. A Home-Box assisted Content Delivery Network System is hence proposed, with two innovative features: (1) a Network- & Context-Aware Anycast-based Server/Peer selection mechanism and (2) an Online Popularity-based Video Caching Strategy. It permits a whole new way of delivering Media contents to End-Users, by combining through a collaborative caching mechanism P2P features with a CDN approach. One PhD student and one Research Engineer have worked on it, in strong collaboration with one SME, the Technological

and Educational Institute of Crete (TEIC) and Portugal Telecom Inovacao. Details are provided in Chapter 4;

Additionaly, we also made an important amount of complementary contributions in the other parts of the system in order to provide a fully integrated solution capable of achieving the related objectives, namely:

- A Cooperative Service Management for Emerging Media Applications, useful for Service and Content Providers. This contribution is detailed in Annex A;
- A Content-Aware Network (Network Elements, Functions and Algorithms) & Management System for Virtual Multi-domain Multi-provider Content-Aware Networks, useful for Service and Network Providers. This contribution is detailed in Annex B;
- A Distributed Framework for Edge and In-Network Media Adaptation, useful for all the potential stakeholders. This contribution is detailed in Annex C.

The document concludes in Chapter 5, by bringing potential outcomes of the proposed contributions and the vision for further work and directions.

Bibliography

[1] Cisco Visual Networking Index, 2011. Available: www.cisco.com

[2] "The Future Internet", John Domingue, Alex Galis, Anastasius Gavras, Theodore Zahariadis, and Dave Lambert (Eds.). Springer-Verlag, Berlin, Heidelberg, 2009

[3] Future Media Internet Architecture Think Tank (FMIA-TT), "Future Media Internet Architecture Reference Model", 1. March 2011, http://www.future-internet.eu/uploads/media/FMIA_Reference_Architecture.pdf

[4] V. Jacobson et al., "Networking Named Content", Proc. ACM CoNEXT 2009. Rome, Italy, December 2009

[5] B. Subbiah, Z. Uzmi, "Content aware networking in the Internet: issues and challenges", Proc. IEEE Int. Conf on Communications, 2001, vol.4, no., pp.1310-1315, doi:10.1109/ICC.2001.936912

[6] F.J. Ramón Salguero, "COntent Mediator architecture for content-aware nETworks", 2010, http://www.comet-project.org/

[7] W. K.Chai, Ning Wang, I.Psaras G.Pavlou, C.Wang, G. García de Blas, F.J. Ramon Salguero, Lei Liang, S.Spirou, Andrzej Beben, E. Hadjioannou, "CURLING: Content-Ubiquitous Resolution and Delivery Infrastructure for Next-Generation Services" , IEEE Communications Magazine, March 2011, pp.112-120

[8] J. Choi, J.Han, E. Cho, T. Kwon, and Y. Choi, "A Survey on Content-Oriented Networking for Efficient Content Delivery", IEEE Communications Magazine • March 2011, pp. 121-127

[9] Cheok, Adrian David, "Art and Technology of Entertainment Computing and Communication", ISBN 978-1-84996-136-3. Springer-Verlag London Limited, 2010, pp. 59

[10] H. Koumaras, D. Négru et al., "Media Ecosystems: A Novel Approach for Content-Awareness in Future Networks", Future Internet: Achievements and Promising Technology, Springer Verlag, pp.369-380. May 2011

[11] EU FP7 ALICANTE project: http://www.ict-alicante.eu/

Chapter 2

A "Networked Media Ecosystem" Architecture for Future Internet based on *-Awareness at different layers

This chapter presents the proposed Networked Media Ecosystem Architecture foreseen in the context of Future Internet and based on *-Awareness at different layers. It first highlights the limitations of the current media architectures before summarizing the research challenges outcoming from such limitations. The state-of-the-art solutions from literature are then described and the proposed architectural solution is finally presented. A summary of the contributions linked to the proposed architecture concludes the chapter.

2.1 Current Media Internet Architectures Limitations

The Internet is flooded with provider- and now more and more with also user-generated media content ubiquitously offered over a variety of applications and in tight integration with social media platforms. There is an increasing tendency towards live communications (in text, voice and video), audio/video consumption and user-generated content publication and consumption.

Consequently, there is clearly a necessity to strengthen the role of networked media content value chain actors so as to be able to respond to this increasing demand of media content and services over the Internet. The inherent overall limitations of the current Internet have been repeatedly identified by recent studies such as [1] and [2]. Especially with regard to the global provision of Media Services, many of these weaknesses are quite critical and pose serious obstacles.

Towards this, we identify, for the course of our study, one *Fundamental Limitation (FL)*: *the best-effort behaviour, associated with the content/service - agnostic nature of the Internet network layer and the lack of cooperation between transport and higher layers can no more satisfy the current and future Media Services needs.* A number of aspects stem from this limitation, which include end-to-end QoS/QoE guarantees, content-related processing at network level, mobility and ubiquitous access, security, personalisation and Media Services adaptation according to the consumption context, etc.

In order to overcome this limitation (but also others), extensive research effort is being conducted in Europe, USA, Japan and other countries towards the Future Internet, either in an evolutionary way or from a clean-slate approach. These research studies have all identified the problem of networked Media Services distribution and consumption and exposed this Fundamental Limitation. It is assumed that this FL, from the network side, directly or indirectly impacts all the actors and also induces other specific limitations.

More specifically, the roles of different actors of the networked media value chain (**Figure 1**) and the limitations[1] they are experiencing are:

- ***Limit_1:*** *service/content creation, publication and dissemination limitations experienced by Content Creators (especially non-professional End-Users);*
- ***Limit_2:*** *service/content management, deployment and context-aware adaptation limitations experienced by Content and Service Providers;*
- ***Limit_3:*** *limitations linked to the user environment capabilities to provide Media Service/content ubiquitous access and context-adapted consumption, experienced by End-Users;*
- ***Limit_4:*** *the Internet best-effort, content-agnostic nature, associated with the lack of cooperation between transport and higher layers, determines the Fundamental Limitation abovementioned and experienced by Network Providers in terms of adaptation of their connectivity services and resources to higher layer services/ applications and induces:*
    - ***Limit_4.1:*** *limited Network Providers capabilities in assuring an efficient and flexible network management, including content recognition, multi-domain collaboration for E2E capabilities, monitoring, traffic handling and quality (QoS/QoE) assurance;*
    - ***Limit_4.2:*** *limited Service Providers/Content Providers capabilities in providing efficient service/content management, deployment and adaptation according to the network status;*
- ***Limit_5****: business model limitations besides technical ones, linked to the insufficient cooperation in terms of management and control between the actors.*

2.2 Future Media Internet Architecture Research Challenges

An approach to overcome the identified limitations, presented in Section 2.1 above, is the creation of a so-called "Networked Media Ecosystem", allowing the integrated management of Media Services in a Future Internet-oriented context. A Networked Media Ecosystem is a common environment, aggregating the involved actors (Content Creators (CCs), Service Providers (SPs), Content Providers (CPs), Network Providers (NPs), End-Users (EUs)) and engaging them in a novel synergy paradigm. Within the Ecosystem, managed creation, publication and delivery of both professional and non-professional generated Media Services are achieved, via the combined exploitation of both centralised and decentralised

[1]A deeper analysis of the current limitations, specific by actor, can be found in [3].

mechanisms. More than that, due to Network-Awareness at service level and Content-Awareness at network level incorporated in the proposed Future Media Internet architecture, more optimised management of services can be achieved at higher layers, together with improved resources management at network layer. This leads to a win-win approach for all the actors of the value chain, thus alleviating their profound limitations and opening them the way to new unforeseen business opportunities.

However, reaching this target implies facing and meeting a number of research challenges. The following sub-sections will detail these challenges and make references to existing research work dealing with them. Each Research Challenge is associated with an identifier (i.e., RC_x) to facilitate future reference. A brief description on how our solution addresses each challenge is also included.

2.2.1 RC_1: Flexible and Integrated Content/Service Management

Limitations addressed: ***Limit_1, Limit_2, Limit_4.2***.

Future media networks foresee flexible content and service management with Context- and Network-aware features, for efficient manipulation and handling [2][5]. These features should be based on relevant (through appropriate filtering) real-time monitoring data not only from the service end-points, but also from other relevant elements involved in the media delivery chain, including the network itself. The different aspects being explored in research targeting this challenge are:

- **Context–Aware Service Management**: this aims at providing users with personalized services that match their context and situation [6][7]. Context-aware service platforms consider the user context at certain steps such as service composition [8] and service delivery [9]. A current on-going EU research project C-CAST [10] is working towards a context-aware communication framework for mobile multicast services with models and ontologies, as well as a reasoning mechanism. We also explore this aspect through models and ontologies but designed a context–aware service management framework that extends the use of user context for uni-/multicast service personalization, including composition;
- **Network–Aware Service Management:** this consists in managing services while being aware of the underlying network in order to dynamically adapt their behaviour according to changing network conditions. This also implies the necessity for open, universal and secure interfaces for the provision of network status in a controlled manner, from the Network Management System to the overlay Applications, while preserving the independence, privacy and security of NP itself (i.e., filtered conforming to the NP policy). Some of the Network–aware systems provide partial caching [11] to adapt to changing network conditions for delivery, while others employ distributed–parallel storage [12]. In our approach, the design is based on continuous network monitoring and provision of certain network metrics to Service Provider in order to facilitate network-aware service management and adaptation;

- **Network-assisted peer-to-peer:** the functionality of P2P overlays can be significantly optimised by exploiting network-side information. Some works such as ALTO [13], the Oracle-based ISP and P2P collaboration [14] and the ISP-driven informed path selection [15] have been carried out towards network-aware P2P delivery network. Hybrid solutions are also considered in [16]. We also provide a Network-Awareness mechanism for leveraging P2P operation, including network-derived Network Distance calculation for best peer selection;
- **End-to-end, cross-layer services monitoring:** it facilitates integrated cross-layer, end-to-end service monitoring at all points of the delivery chain for achieving service assurance, fault detection and facilitate cross-layer adaptation decisions. The recent FP7/ICT ADAMANTIUM project [17] worked towards achieving cross layer monitoring and adaptation for media streams. Relying on ADAMANTIUM results, we provide a distributed cross-layer monitoring framework for Media Services;
- **Facilities for the creation and addition of new services and functionalities:** this consists in the capability for on-demand activation of a new service, benefiting from provisioned network functionalities and protocols. Future media platforms should incorporate such flexibility. The IP Multimedia Subsystem [18] offers a common platform to reduce the rollout time for new Media Services. Thanks to XML-based interfaces, Service Oriented Architectures (SOA) facilitates service deployment and makes them OS- and language- independent. Our solution is providing SP/CPs with an SOA-based open platform that enables an easy service integration and deployment.

2.2.2 RC_2: Enhanced User Experience

*Limitations addressed: **Limit_1, Limit_3**.*

Meeting End-Users' needs, interests and tastes is becoming a key competitive factor for Content and Service Providers. Despite the progress that has been made in the last decade, providing End-Users with a strong and compelling experience remains a challenging issue. The following highlights the key features, foreseen as essential to improve user experience, for which intensive research is ongoing:

- **Context-Awareness** is necessary in order to offer mechanisms that enable the service provisioning process to be aware of any contextual information considered as relevant to the interaction between the End-User and the service [19]. Since the End-User experience is highly affected by his/her surrounding context, the latter should then be collected or inferred from the previous users' preferences and activities [20], modelled [21][22][23] and managed in a way that leads to maximize user satisfaction. The context-aware feature is even more important today with the multiplicity of EU's possible terminals, locations, and preferences, from one side, and the diversity of accessible Media Services, on the other. Being a key solution for service convergence and integration, Context-Awareness has been considered in several standardisation works such as NGN [24], IMS [18][25], NGSON [26], MPEG-7 and MPEG-21 [27] and addressed by many research projects such as Mobilife [28], PERSIST [29] and COAST [30]. So far, by always targeting a specific type of application and context of use, none has reached to a full generic solution capable of meeting all End-Users diverse needs.

Our solution tries to do so by also considering these approaches and proposing an integrated service-independent solution for Context-Awareness in maximum compliancy with existing standardisation work;

- **Ubiquity** is a major trend in today computing. The End-User wants to access contents and services anytime, anywhere and through any device or network. In addition to the personalization feature, seamless **mobility** is very important to ensure good user experience. Many standardization activities have been carried out to provide mechanisms that enable network mobility such as Mobile IP (MIP) [31], MIH IEEE 802.21 [32], IMS-Cont [33], etc. From a user-centric view, Mobilife [34] proposes a context-aware driven mobility and PERIMETER [35] a QoE-driven mobility. We do not directly tackle mobility aspects but are compliant to rely on those existing solutions;
- **QoE evaluation:** it is very important to get real-time feedback about the user experience in order to keep the latter at a satisfying level. The assessment of subjective QoE remains a challenging issue and still attracts much interest from the research community. The study carried in [36] gives a survey and a classification of basic quality assessment methods and metrics and the work [37] presents a method that uses artificial neural networks (NNs) to automatically quantify the quality of video flow. We also aim at client-side QoE assessment towards fulfilling QoE requirements for End-Users. Towards this, we propose a solution whose concept and approach rely on [37];
- The advent of **User Generated Content** is reshaping the way End-Users experience services [38]. End-Users are allowed to be not only consumers but also producers, i.e., prosumers. It is obvious that in this context, End-Users need efficient tools to easily **create** and **self-publish** new contents and services. Our solution exploits existing technologies on service creation and composition and proposes an integrated framework for user-friendly deployment of user-generated composite Media Services.

2.2.3 RC_3: Content Awareness at Network level

*Limitation addressed: **Limit_4**.*

Content-aware networking (CAN) is a new paradigm in network engineering, aiming to evolve or even restructures the classic TCP/IP and OSI stack concepts. Being a rather new concept, there are no significant standardisation results yet. The Content-Awareness (CA) term refers to the paradigm shift from the current, content-oblivious architecture to a network-aware of the transported content. Content-Awareness features the following functions:

- **Content-type or content-objects recognition** at network level by the network elements themselves, either automatically or using explicitly authorised signalling by higher layers, [39][40][41]. We adopt a flexible set of content-type recognition methods, based on Data Plane and Control Plane methods;
- **Service classification and differentiation:** this consists in the idea being applied in all well-known technologies related to QoS (DiffServ, MPLS, IntServ, etc.). The CA feature we propose enhances and enriches this by giving possibility not only to provide QoS according to content-type and associated policy information, but also to add some other CA related processing specific to services;

- **Content-Aware Caching,** (c.f. e.g., [39]), is used in order to make availablethe content objects from within the network elements (to achieve shorter paths from the user to the content source). Instead of applying the classic in-network caching approach, weare proposing CA caching in the service end-points (user media gateways);
- **Content-Aware Processing:** it refers to filtering, policing, scheduling, of the traffic flows by the network elements, in an efficient way and with guarantees, but still providing a scalable solution. Policing and scheduling has been treated in many studies and applied into projects [42][43][44][45][46][6]. Our solution relies on existing work and applies CA processing according to the content-type with different appropriate policies;
- **Content-Aware Routing and forwarding:** it consists in the location/identity decoupling, especially in the routing and forwarding process, which is related to CA approach, making the networks' service more flexible for content oriented communications. New trends exist to solve the features of content/name based routing and forwarding [39][40][6][49][50][51][7][8]. We apply the CA based forwarding and partially CA routing.

2.2.4 RC_4: Multi-domain network resource management, virtualization and security

*Limitation addressed: **Limit_4**.*

Multi-domain network resource and QoS management is still a hot topic given the need for synergic management of multi-domain networks to offer E2E QoS. On the other hand, independency of each network domain is always a prerequisite. The following challenges exist:

- **Multi-domain QoS support**: the MESCAL [54][42], ENTHRONE [43], AGAVE [44][45] EU projects are only a few examples of architectures engineering the Internet to support QoS across multiple domains by defining end-to-end QoS as a composition of local QoS in each of the domains, negotiated between peer Network Providers. However, they do not specifically address media content, while our solution for inter-domain peering is considering the Content-Awareness feature;
- **Dynamic assignment, provisioning and interfacing** of customizable multi-domain network services to upper layers (e.g., SPs): this challenge is tackled in [43][44][55]. However the aforementioned works do not address the cross-layer optimisation between the network layer and upper layers. We propose this based on CAN/NAA (Content-Aware Networking – Network-Aware Applications) concepts.

Virtualisation is considered as a key feature of future network infrastructures, as it enables increased flexibility and collaboration capabilities among Network and Service Providers. Related to virtualization, the following challenges are highlighted:

- **Creation of virtual networks while making abstraction of a certain subset of network resources** (link bandwidth, element processing power, etc.). Several parallel slices can be considered, as well as identification of technology-dependent (e.g., MPLS, carrier Ethernet, etc.) and technology-independent mechanisms for virtualisation. Virtualisation concepts [56][57][58][59][46] based on overlays have been proposed as a powerful method to make the network environment more flexible and customisable. Parallel Internet planes [44][45] have been proposed to split the network into customisable logical networks. However, these solutions did not consider offering advanced network services such as Content/Service-Awareness at virtual network level. Our solution offers logically isolated virtual and content-aware networks (VNets) customisable for different types of content and requirements;

- **On-demand provisioning of virtual network services** (e.g., security, Content-Awareness) to be offered to upper layers virtual networks, e.g, in [46], by defining a VNet Provider and Operator dealing with these. Our solution applies similar concepts, having additional capabilities related to Content-Awareness;
- **Enhanced network management schemes** which support VNet planning advertising/discovery/offering, negotiation, provisioning, operation (installation, modification, manipulation, monitoring, termination) performed at virtual network layer, while cooperating with the traditional management of the IP network layer [44][45][46][7]. Our solution relies on existing work;
- Support for **inter-domain peering** and creation of VNets across multiple independent physical network domains, conforming to certain SLA/SLSs, while preserving each domain's resource management independency [44][45][46][7]. In [59], it is shown how to use virtualisation to provide inter-domain QoS-enabled routing. Our solution relies on it to develop an inter-domain virtual topology including several domains in virtual CANs;
- Support of **unicast and multicast services** on top of the virtual networks. The CURLING [7] architecture supports content-centric overlay networks using a multicast-style receiver-driven service model, but does not address content adaptation, mapping to native IP multicast, or QoS. In [60], support for multicast streams adapted to each terminal's needs is proposed, by encoding media in multiple SVC layers, and defining independent multicast trees for each layer, but it only supports overlay multicast. Our solution caters for hybrid native IP/overlay services for multicast distribution in order to maximise resource usage.

Network Security is a hot topic in Future Internet. Among the challenges, one can be mentioned more directly related to media aspects:

- **Support for DDoS attack detection, traceback and mitigation services** [61][62][63][64] combined with network layer packet filtering. Current DDoS (Distributed Denial of Service) mitigation techniques do not cope with network services that are being offered on top of virtual networks. Moreover, existing solutions tend to consider, on one side, that there is only one victim and, on the other, that any other Internet connected equipment is a possible attacker. Our solution considers the provisioned virtual network as a whole, and exploits estimations for aggregate network usage, based on statistics from network end-points, as a basis for detecting traffic anomalies.

2.2.5 RC_5: Edge and in-network media adaptation

Limitations addressed: Limit_2, Limit_3, Limit_4, Limit_5.

Content adaptation being an effective and attractive solution to the problem of mismatch among content format, network conditions, device capabilities, and user's preferences, has become an ongoing hot topic in research. The adaptation is decided and triggered, as a policy enforcement action, by a user context- and/or network-aware Service Management framework, as aforementioned, so that media delivery matches user context and network status. The aim is to achieve both service personalisation and QoS/QoE assurance by exploiting the following possible features for adaptation:

- **Server-side adaptation,** for optimisation of one-to-many single-stream services, such as multicast streams, or for multiple content encoding for simultaneous delivery to various user contexts (e.g., fixed, mobile, Standard/High Definition (SD/HD), etc.). Our solution performs server-side adaptation for both unicast and multicast services;
- **In-network adaptation,** where media streams are processed by the network elements themselves. This approach is best suited for multicast traffic, where different delivery paths may have different capacities. To avoid imposing significant processing overhead at the network elements, Scalable Video Coding (SVC) schemes [66][67] are best suited for this purpose. SVC facilitates edge routers or in-network elements to process or drop packets depending on how many layers are allowed in the target network. A complexity analysis of SVC for IPTV services is presented in [65]. Our solution is making use of SVC at the transport level and applies in-network adaptation by dropping layers in case the management requires it;
- **Client-side adaptation**, in the case where media streams are to be re-transmitted or re-distributed in a local manner (e.g., within a home network, to a location unreachable directly by the SP/CP). Our solution proposes client-side adaptation inside the residential gateway, for context-aware, terminal-dependent adaptation.

2.2.6 RC_6: Open and efficient collaboration among actors

Limitation addressed: Limit_5.

The collaboration of different involved actors (CCs, SPs, CPs, NPs, EUs) is essential for building an ecosystem for networked media. The lack of collaboration in current architectures limits the scope of features by the actors. Existing work has identified several research challenges for:

- **Efficient collaboration related to management**: MANA Group [1] evaluated as challenges several issues, such as: the facilities to support QoS and SLAs, self-management, facilities for the large scale provisioning and deployment of both services and management; support for higher integration between services and networks, capability for activating a new service on-demand, network functionality, or protocol (i.e., addressing the ossification bottleneck), etc. We answer partially to these by developing a powerful management system;
- **Control Plane evolution:** related challenges exist in the current Internet such as the lack of flexibility in control architecture [68][69]. The "hour-glass" IP model does not account for this evolution of the control functionality when considered as part of the

design model. In addition, the lack of unified architecture of the IP Control Plane is growing. Even though the original IP Data Plane is simple, its associated control components added during years are numerous and their interactions are more and more complex [70][71]. Our solution contributes to developing a lightweight Control Plane, managing both virtual networks and underlying network infrastructure.

Many studies have been devoted to certain number of research challenges directly linked to the improvement of Networked Media systems. In the following sub-section, we will highlight some of them.

2.3 Related Solutions for Future Media Internet Architectures

This subsection overviews current and past research effort, which addresses the research challenges previously mentioned with regards to future media networks. We have selected architectural reference models that have been proposed by large Think-Tank groups, such as FIArch and FMI-TT, or by important research projects. The list of selected studies to be shortly analysed below is not exhaustive. Correlating their scope with ours, the selected solutions:

- Are (partially) media and content oriented, including end-to-end QoS;
- Consider multi-provider, multi-domain, multi-technology architectures;
- Cover (partially) the integrated management of both high-level services and networking resources.

The objective of this subsection is to identify the scope and limitations of the proposed solutions in order to clarify our proposal's complementarities and/or progress with respect to them.

2.3.1 FIArch and FMIA-TT

The Future Internet is an emerging area where already a lot of on-going work is done. Through its Internet Architecture Task Force, FIArch has highlighted the main limitations of the current Internet architecture, which the Future Internet design should take into account [72]. A detailed up-to-date FIArch work can be found in [118]. New Internet architecture models are proposed in [46][47][48][49][50][73][74]. FIArch [4] has started working at architectural level for the improvement of the whole Internet architecture and principles, incorporating, among others, media content aspects to which our proposal has one of the primary roles [75].

The Future Internet is expected to be a communication and delivery ecosystem. As an extension of the FIArch (today regrouped together), the Future Media Internet Architecture – Think Tank group (FMIA-TT) aims to specify a reference model of a "Future Media Internet Architecture" which covers delivery, adaptation/enrichment and consumption of media within the Future Internet ecosystem. Consequently, the FMIA-TT has defined [5] a high-level FMI network architecture based on four macro-layers (or strata) (as depicted by **Figure 2**):

- ***Applications Overlay (AO)***: it includes applications which may use different services and the information delivered by the Information Overlay (IO) and the media/content themselves;
- ***Information Overlay (IO)***: it comprises intelligent nodes and servers with knowledge of the content/Web-service location/caching and the network instantiation/conditions (the nodes can vary from P2P peers, secure routers or even Data Centres);
- ***Distributed Content/Service-Aware Overlay (DCSAO)***: it contains the CAN nodes which filter content and Web services;
- ***Service/Network Provider Infrastructure (SNI)***: it can be seen as the traditional layer of services offered by the ISPs. The users can be providers and/or consumers ("prosumers") of the services offered by this layer.

More details on FMIA-TT reference model can be found in [5].

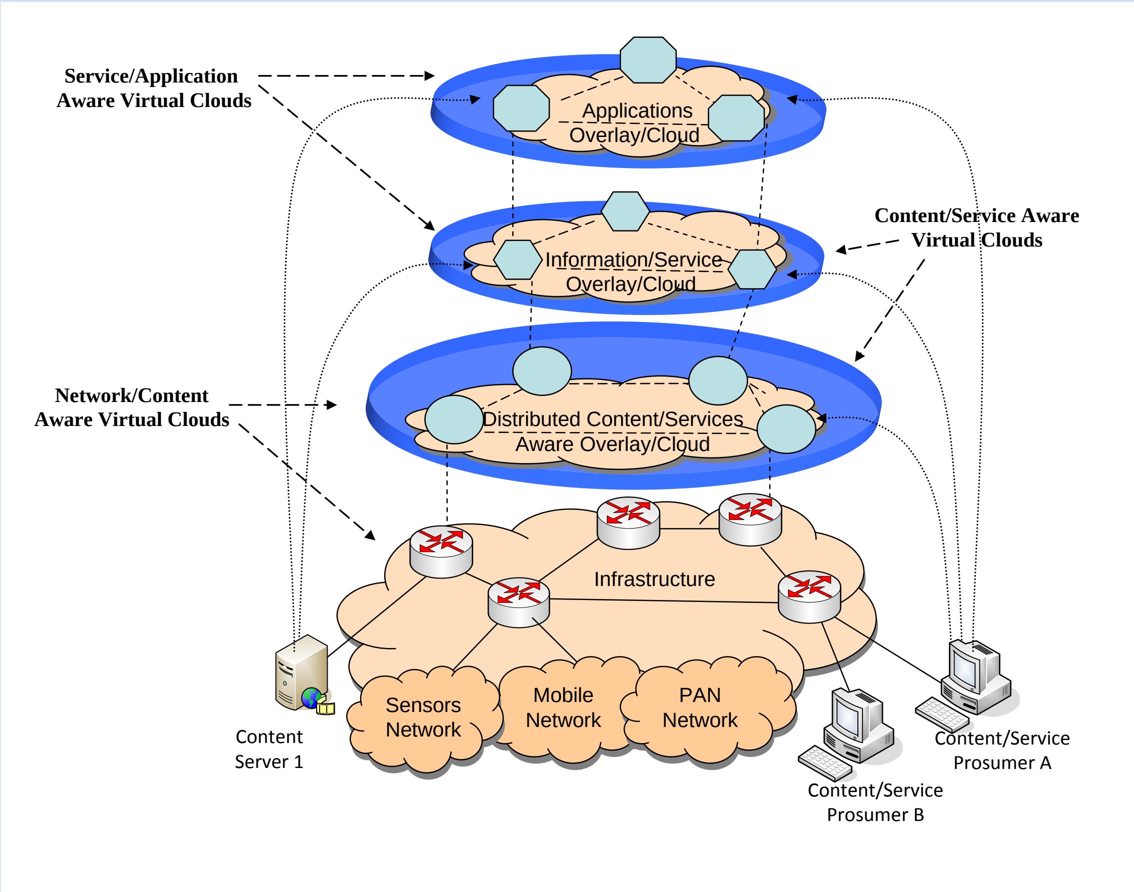


Figure 2: FMIA-TT Reference Model [5]

The architecture we propose, splitted in environments, is compliant with the FMIA-TT one.

2.3.2 Related Research Projects

Few other solutions towards similar objectives of Future Media Internet architectures are also presented below.

The FP6 project **MESCAL** "Management of End-to-end Quality of Service Across the Internet at Large" [54][42] proposed an evolutionary, scalable, incrementally architecture, enabling flexible deployment and delivery of inter-domain QoS across the Internet at large. The MESCAL business model actors are: Service Providers (SPs), IP Network Providers (INPs), Physical Connectivity Providers (PCPs) and Customers. MESCAL developed a generic, multi-domain, multi-service functional architecture, and a complex management system mainly focused on resource management and traffic engineering (offline and online) intra and inter-domain. However, several MESCAL limitations exist with respect to our proposed solution: the End-Users cannot act as Content Providers; it is focused mainly on networking aspects and does not consider the service; it does not have a multimedia orientation as a main design direction. Content-aware networking, P2P mode, HBs are missing. While our solution may use the MESCAL concepts of QoS classes (local, extended, meta-QC) in a multi-domain environment, it also brings the additional abovementioned features missing in MESCAL.

The FP6 project **ENTHRONE** "End-to-End QoS through Integrated Management of Content, Networks and Terminals" [43] proposed an evolutionary complex architecture on top of IP, to cover an entire Audio/Video (A/V) service distribution chain, including content generation and protection, distribution across QoS-enabled heterogeneous networks and delivery of content at user terminals. ENTHRONE targeted primarily multimedia distribution services. Although it is focused on high-level Media Services and also some networking aspects, ENTHRONE User and Service environment have limited capabilities in terms of complex services management, composition and mobility. Content-aware networking, P2P mode, HBs are missing. Our proposed architecture addresses such aspects.

The FP6 project **AGAVE** "A liGhtweight Approach for Viable End-to-end IP-based QoS Services" [44], tries to solve the end-to-end provisioning of QoS-aware services over multi-domain IP networks. The business model defines the Service Provider (SP) and the IP Network Provider (INP) business roles. The architecture is based on the novel concept of Network Planes, allowing multiple INPs to provide Parallel Internets (PI) tailored to E2E service requirements. However, AGAVE lacks a complete chain of complex services management: definition, creation, offering, exploitation, composition, etc. It also does not consider home networking and content-aware aspects. Our proposed architecture benefits from the AGAVE concepts of PIs, by offering the VCAN as enhanced equivalent of Network Planes, but in the framework of a more complete architecture, of the proposed Media Ecosystem.

The FP7 project **4WARD** "Architecture and design for the future Internet" [46], is a large research project clearly oriented towards FI. It proposes new Architecture Concepts and Principles (NewACP) based on a plurality and multitude of network architectures: the best network for each task, each device, each customer, and each technology. Networks coexist and complement each other, each of them addressing individual requirements such as mobility, QoS, security, resilience, wireless transport and Energy-Awareness. The business players are: Physical Infrastructure Provider (PIP), Virtual Network Provider (VNP) (assembling virtual resources from one or multiple PIPs into a virtual topology), Virtual Network Operator (VNO) (installation/operation of a VNet over the virtual topology provided by the VNP for a tailored connectivity service), Service Providers (SPs) (use the virtual network as a support for their services - these can be value-added services and then SPs act as application service providers, or transport services and then SPs act as network service providers). Full network virtualisation is considered in 4WARD to solve the interoperability and is considered a main concept for a clean slate FI approach. Our proposed architecture benefits from the virtualisation aspects investigated in 4WARD, however it aims at an evolutionary, backwards-compatible approach rather than a clean-slate one, in order to maximise adoption possibilities and to accelerate market penetration.

The FP7 project **PSIRP** "Publish-Subscribe Internet Routing Paradigm" [49][50], claims to be a clean slate FI approach. It aims to develop, implement and validate an Internet working architecture based on publish-subscribe paradigm, as a promising approach to solve many of the biggest challenges of the current Internet (alternative to the commonly used Send-Receive paradigm). The business model is composed of: publishers, subscribers, and a network of brokers. As a follow-on, the EU FP7 project **PURSUIT** "Publish-Subscribe Internet Technologies" [51], further explores and expand PSIRP's vision, targeting a more complete architecture and protocol suite, more performing and scalable. The PSIRP is revisited to produce and evaluate alternative designs, as well as to expand system dimension (hundreds of nodes), dissemination and exploitation. PURSUIT architecture maintains the PSIRP flat-label-based information identification. Every information item in PURSUIT is associated with – at least – one scope, and the information organization follows the same PSIRP principles. Our proposed solution, although it does not fully adopt the publish-subscribe paradigm, introduces novel media discovery and distribution mechanisms beyond traditional ones.

The FP7 project **COMET** "COntent Mediator architecture for content-aware nETworks" [47][48] aims to provide a unified interface for content access whatever the content characteristics are: temporal nature (pre-recorded or live), physical location (centralised or distributed), interactivity requirements (elastic or real-time), or any other relevant features. It also aims to apply the most appropriate end-to-end transport strategy: by mapping the content according to its requirements and user preferences to the appropriate network resources; best quality of experience for End-Users; it supports unicast, anycast and multicast. All these are achieved while preserving network availability and structural resilience, as key factors in perceived QoE. The business entities are: Content Consumers (CC), Content Providers (CP), COMET-capable ISPs and carriers. COMET aims to be a flexible framework to accommodate several possible current or future content-related business models. COMET distinguishes two kinds of scenarios: free content access and charged content access. It is identified that overlapping domains exist between the COMET proposal and ours. However, the COMET business model is only partially sufficient for our needs; it does not consider fully the cooperation between network overlay and network resources, but is focused mainly on mediation activities. Also there is no complete chain in terms of complex services management and adaptation and COMET uses a Content-Object oriented approach (CON [76]) whereas we are more on a Content-Type oriented one (CAN), less revolutionary but more incrementally deployable.

In [77], a revolutionary solution is proposed as *Content-Centric Networking (**CCN**).* The idea stems from the fact that the IP networks are increasingly used for content distribution and retrieval, while networking technology are still based of connections between hosts. CCN replaces the traditional "where" paradigm used for IP routing with "what"- identifying the content by taking the content as a primitive – decoupling location from identity, security and access, and retrieving content by name. Using new approaches to routing named content, derived heavily from IP, the study claims that one can achieve simultaneously scalability, security and performance improvements. While not applying full CCN concepts, our solution is capable to be incrementally developed towards a CCN one, given its Content-Awareness feature incorporated in the edge routers.

In [7], a mixed content centric solution is proposed. The ***CURLING*** architecture, "*Content-Ubiquitous Resolution and Delivery Infrastructure for Next Generation Services*", aims at efficiently diffusing media content of massive scale. It entails a holistic approach, supporting content publication, resolution and content delivery, and provides to both CPs and customers high flexibility in expressing their location preferences, when publishing and requesting content, respectively, through *scoping* and *filtering* functions. Content manipulation operations can be driven by factors as business relationships between ISPs, local ISP policies, and specific CP and customer preferences. Content resolution is also coupled with optimized content routing techniques that enable efficient unicast and multicast- based content delivery across the global Internet. Our proposal and CURLING's one are complementary. We are less content-centric in the Control Plane, but more powerful in assuring efficient media flow QoS enabled transport based on Content-Awareness and in completeness of the service management framework.

The FP7 project ***OCEAN*** "Open ContEnt Aware Networks", [60], designs a new open content delivery framework that optimizes the overall QoE to End-Users by caching content closer to the user than traditional CDNs do and by deploying network-controlled, scalable and adaptive content delivery technique. OCEAN aims to find solutions to the imminent problem of multimedia content traffic clogging up the future aggregation networks, when the offering of online video of high quality over the Open Internet continues to increase. OCEAN builds innovative self-learning caching algorithms that meet the specifics of the highly unpredictable location and time-dependent consumption patterns and dynamically adapt to the rising popularity of future delivery services. Media-aware congestion control mechanisms based on slight, but controlled quality degradation is suggested rather than blocking of user requests. OCEAN and our approach can bedifferent in some aspects and complementary in others, in the sense that the first addresses Content-Awareness from the caching point of view, whereas the second mostly studies traffic differentiation issues. As well, we propose a caching strategy but in specific equipments.

The FP7 project ***P2P-Next*** [78] aims to build a next generation Peer-to-Peer (P2P) content delivery platform. The objectives of P2P-Next are the distribution of radio and television programmes, movies, music, ring tones, games, and various data applications to the general public via a variety of dedicated networks and special End-User terminals. P2P-Next uses the BitTorrent protocol to deliver SVC encoded content to the End-Users. Furthermore, P2P-Next provides adaptation at the client side via layer switching at instantaneous decoding refresh (IDR) frames. Our proposed solution will be protocol independent, allowing the use of P2P protocols, unicast or multicast protocols (e.g., HTTP, RTP). Further, our approach not only offers adaptation (i.e., transcoding, rewriting) at the client side but also in the network by dropping SVC layers within the network nodes. Considering P2P case, our solution relies on P2P-Next open source one.

Much work has been devoted to certain number of research challenges directly linked to the improvement of Networked Media systems. Our proposal inspires and relies on existing work from research projects and studies, when possible, and tries to be as much compatible with existing standards as possible, when not directly adopting them (e.g., MPEG DASH, SVC, RTP, etc.).

The following section describes our proposal of a Networked Media Ecosystem (NetME) architecture towards Future Internet.

2.4 The "Networked Media Ecosystem" (NetME) Architecture Proposal

The shaping and high-level definition of the proposed Networked Media Ecosystem (NetME) architecture, towards Future Internet, is derived from the actual context encompassing the actors' limitations, the associated research challenges and the current state-of-the art, as identified in previous sections.

This architectural proposal [75] has been accepted as the main European research project in the Networked Media domain[2]. Its implementation and large-scale evaluation have been performed within this project.

2.4.1 High Level Architecture Description

The principal objective is to address the limitations linked to the actors of the networked media content value chain and thus present an architecture working around their environments and assuring maximum cooperation, towards the creation of a Networked Media Ecosystem. The solution is a layered architecture, in which layer/environment functions are mapped to the corresponding actors able to accommodate (1) the current and future needs of media content oriented services and (2) the flexible, scalable and efficient usage of network transport resources over heterogeneous networking technologies. In this respect, the new architecture is based on Content-Aware networks and Network-Aware services paradigms, towards the creation of a Networked Media Ecosystem, supporting flexible, efficient and intelligent usage of network resources based on cooperation between existing actors and systems. The architecture comprises a number of environments and layers aiming to improve today's media delivery networks and systems and to enable enhanced QoE and additional services for End-Users.

[2] ALICANTE is a funded EU FP7 IP research project under the Networked Media and Systems programme. It consists of 20 partners. http://www.ict-alicante.eu

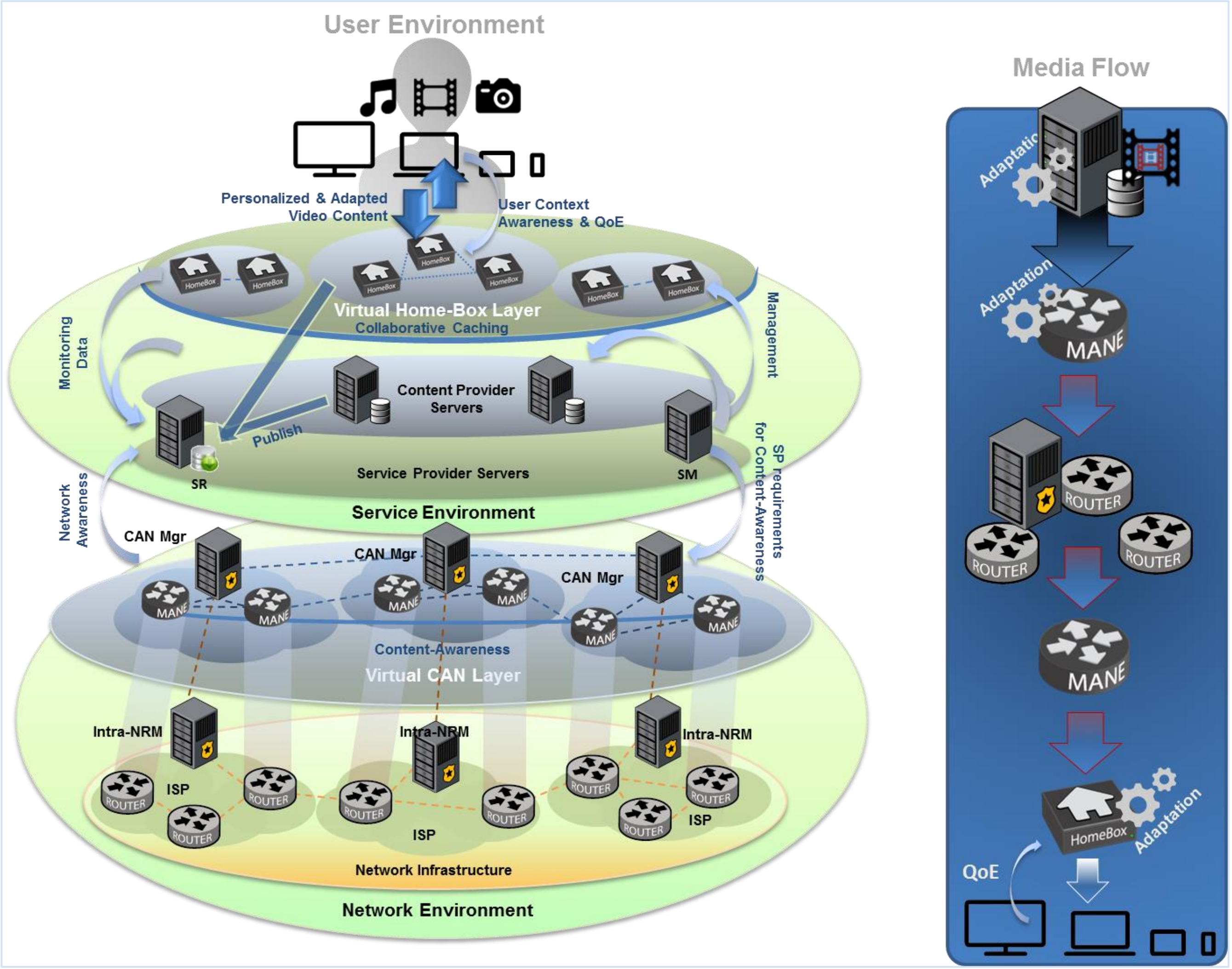


Figure 3: The Proposed High-Level Architecture Overview

Aiming to achieve a managed Networked Media Ecosystem, the proposal, illustrated in **Figure 3**, is broken down to architectural layers and environments, as follows:

- **The User Environment (UE)** allows the End-Users (EUs) to consume and/or generate content and exploit different services delivered by components of the Service Environment. Most of the User Environment functions are to be included in the EU terminal(s), which are connected to the user's residential gateway (Home-Box). The User Environment is the architectural part of the system dedicated to address Research Challenge RC_2 (confronting limitations *Limit_1* and *Limit_3*). Towards this, the User Environment offers to the EU the potential to have several roles, such as Content and Service Consumer, Provider or Manager (i.e., becoming a real "Prosumer"). This is made possible through a generic multi-platform, user-friendly graphical interface permitting the EU to access/deliver/manage any service/content on any device from anywhere and at any time. A User Profile is foreseen, characterizing the static and dynamic parameters of the user and his/her context, in order to be exploited by Service Environment elements for the delivery of context-adapted services (Context-Awareness). For dynamic part, the User Profile relies on Quality of Experience and Quality of Services monitoring;
- **The Service Environment (SE)** offers enriched networked Media Services and manages their whole lifecycle (creation, provisioning, offering, delivery, control). In the Management and Control Plane, the SE receives user context information from the UE and network information from the CAN/Network layer, to achieve a context- and network-aware service provision. In the Data Plane, the SE uses the overlay connectivity services of the CAN layer and is also involved in the process of adapting the services/content according to the EU context. Service composition, security and privacy are other functionalities considered at SE level. End-to-end Service Management is supported in an integrated manner, by the Service/Content Provider infrastructure, as well as evolving capabilities for combination and interoperability to existing platforms (especially standardised ones). The Service Environment is the architectural part of the system dedicated to address Research Challenge RC_1 (confronting limitations *Limit_2*). Towards this, it allows providers to offer enhanced services and content through:
    - A cluster of media servers (*SP/CP servers*);
    - A *Service Registry* functionality containing detailed information on all available services. It maintains updated information from all providers (including End-Users publishing user-generated content), and End-Users rely on it to discover available services and contents;
    - *Service Composition, Service Management and Service Monitoring* features, supporting the entire service lifecycle ;
- **The Network Environment (NE)** (including CAN and Network layers) realizes and offers to upper layers a rich and virtualized networked space, which can be customized and exploited by all business media sectors for delivering networked media content. The CAN & Network layers of the NE are the architectural parts of the system dedicated to address Research Challenges RC_3, RC_4 and parts of RC_6 (confronting limitations *Limit_4, Limit_5*). Toward this, the NE includes:

  - A new **Virtual CAN layer**, offering connectivity services on top of the IP infrastructure by constructing mono- or multi-domain Virtual Content-Aware Networks (VCANs). The VCANs are virtual networks offering enhanced support for packet/flow processing in network nodes. They can improve data delivery via content-aware traffic classification and appropriate processing (filtering, routing, forwarding, QoS-processing, adaptation- dynamic, aggregated or per flow, security, monitoring). The CAN layer is managed by dedicated modules (CAN Managers – CANMngr), one per network domain. The decision to have one CANMngr per network domain assures seamless deployment in real environments;
  - **IP network infrastructure layer**, instantiating the CANs via its Intra domain Network Resource Managers (Intra-NRM) at request of the CAN Manager. The advanced Network- and CAN-layer traffic handling capabilities are supported by enhanced/upgraded network elements: the Media-Aware Network Elements (MANEs) ;
- **The Home-Box (HB) layer** is composed of virtually interconnected Home-Boxes, capable of advanced ways of service/content provisioning. The Home-Box layer is the architectural part of the system, within both User and Service Environment, enabling Context- and Network-Awareness to Service part. It hence addresses the respective challenges (RC_1, RC_2, and parts of RC_6). Toward this:
  - The *"***Home-Box***"* entity is a new media-centric Home Gateway, featuring advanced functionalities, such as native service provision (to other HBs), service adaptation and redistribution (to the Home network and the associated User terminals), user/service mobility and security;
  - The **virtual HB layer** is an overlay composed of distributed HBs offering a flexible logical infrastructure, configurable in hierarchical unicast/multicast distribution mode, in a distributed mode (e.g., P2P) or in a combination of both. The virtual HB layer promotes distribution of both Provider- and User-Generated services among HBs in a flexible and optimised way;
- **Cross-layer Monitoring system,** spanning all environments/layers, supports end-to-end monitoring functions at all levels, for services and for resources. This system is not explicitly represented in **Figure 3**, but its functions are part of the managers and their interactions. The Cross-layer Monitoring system is providing the inputs to the Media Adaptation system and all the relevant information to perform the *-Awareness features;
- **Media Adaptation system**. The adaptation process itself can occur either at the Content Server side (at the SP/CP premises), inside the HB, in a network element (targeted by CAN decision) or in a combination of places. It can be launched on a unicast service but also on a multicast one, through the use of SVC. It is triggered by an Adaptation Decision Taking Framework (ADTF) based on monitoring system information and inter-actor established agreements (SLAs). The Media Adaptation is

the architectural part of the system addressing Research Challenge RC_5 (confronting limitations *Limit_2, Limit_3, Limit_4, Limit_5)*.

The aforementioned Environments comprise well-defined interfaces enabling the exchange of content- and network-based information between the Network and Service Environment – supported by a flexible and distributed Cross-layer Monitoring system – and, thus, realising Content-Aware Networks and Context+Network-Aware services. This *-Awareness principle is exploited in two ways: (1) by constructing multi-domain, multi-provider Virtual Content-Aware Networks having customisable content-aware behaviour (QoS provisioning, security, forwarding/routing processing) and (2) by a distributed Adaptation Framework, which performs edge- and in-network service and content adaptation in order to optimize the offered QoE, towards the best possible End-User's Media experience.

More details on the architecture and concept can be found in [3].

2.4.2 NetME System Architecture Design

From a design perspective, the NetME overall system architecture has been organized as depicted in **Figure 4**. Inline with the overall architecture presented in 2.4.1, it comprises five subsystems and associated components (detailed in the following paragraphs). It must be noted that all the subsystems have a monitoring module, interacting with subsystem's inner modules, whereas the adaptation system is distributed, each subsystem's adaptation module is interacting with the others.

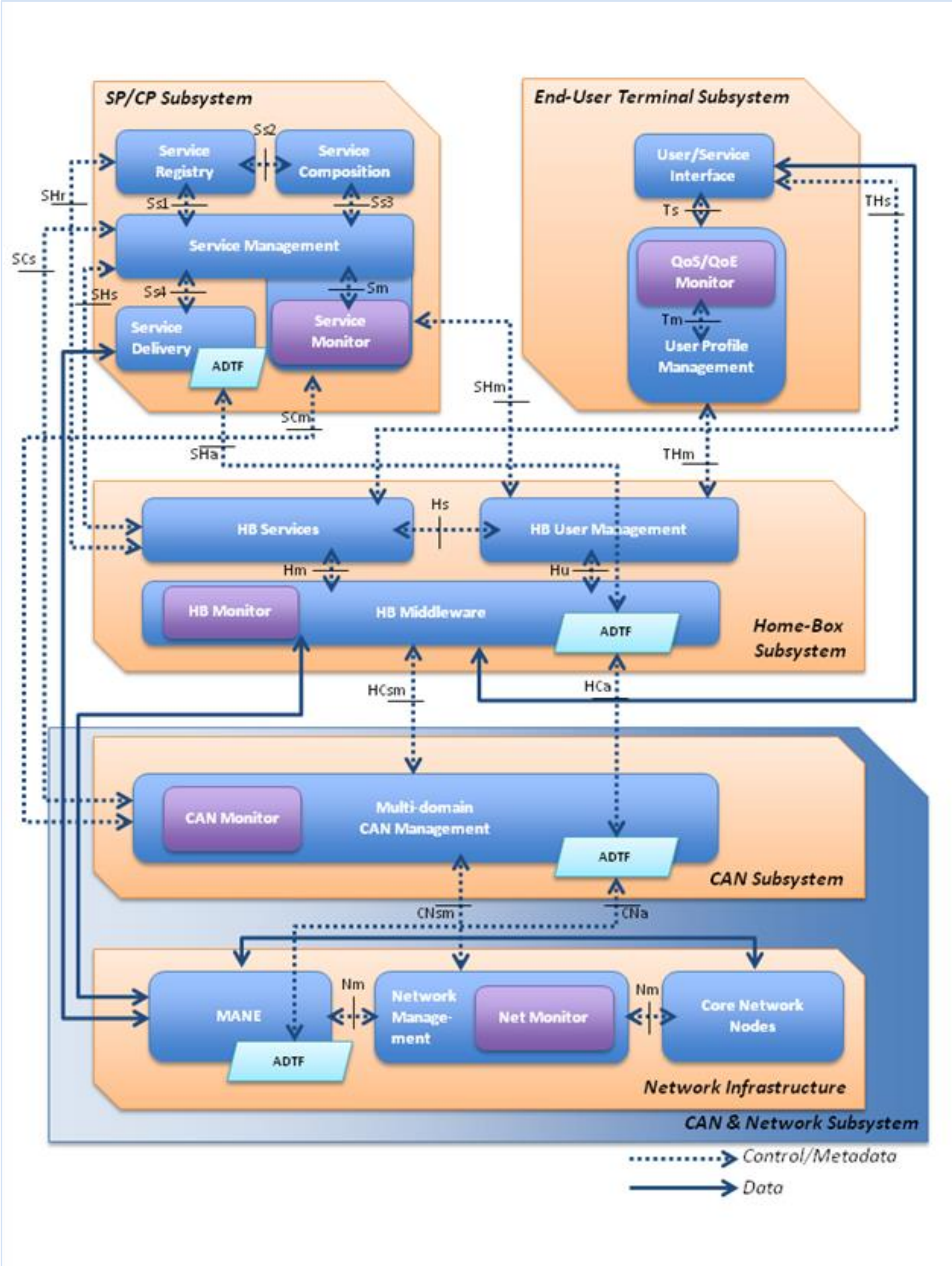


Figure 4 : NetME System Architecture Design

2.4.2.1 The End-User Terminal Subsystem

As the essential part of the User Environment, the End-User Terminal (EUT) subsystem presents specific functionalities and interacts closely and seamlessly with the Home-Box entity to provide End-Users with enhanced Media Services. In detail, the following components compose the EUT,enabling advanced functionalities, such as **Context-Awareness**:

- **The User Profile:** it is a functional bloc in charge of collecting and managing data that characterize both the user and his operational context, composed of static and dynamic information. It has two functions: 1) to gather and manage context information in order to keep the User Profile instance up-to-date and accurate, and 2) to provide the relevant part of the User Profile in order to be available (to and via the Home-Box) to other entities, allowing them, on one hand, to invoke and compose suitable services, and on the other hand, to adapt in real-time their behaviour according to the User Profile (Context-Awareness);
- **The QoS/QoE monitor:** the purpose of the QoS/QoE Monitor functions at the End-User Terminal is twofold: to quantify the ongoing subjective experience of the EU and to trigger alarms for adaptation procedures, when necessary. The QoS/QoE Monitor shall collect both network/transport and also user-level (i.e., perceived quality) metrics, in order to monitor both network-level QoS (i.e., impairments at incoming media flows) and also perceived QoS/QoE. A lightweight psychometric model is used to estimate the perceived QoS/QoE from the measured network metrics. Different monitored parameters are sent to the User Profile database through update messages;
- **The User/Service Interface**: it constitutes the system front-end, as offered to the End-User for both service consumption and publication. The User/Service Interface is terminal-independent, and auto-adjusts according to terminal capabilities (e.g., screen resolution).

The EUT subsystem brings innovation in the field of Context-Awareness, by proposing, through the User Profile and QoS/QoE monitor modules and in conjunction with the Home-Box Subsystem (where more details are given), a **"Context-Aware Framework for Media Services Provisionning"**. This contribution is presented in Chapter 3.

2.4.2.2 The Service/Content Provider Subsystem

The SP/CP Subsystem's purpose is to provide the essential functions enabling the provisioning of Media Services coming from various sources, to register and publish them in order to be searchable by the End-Users, to plan, negotiate and request network resources for the delivered content (input for **Content-Awareness** at Network level), to manage the subscription and access to the Media Services and to deliver them to End-Users according to the contracted SLAs and their context (**Context-Awareness**). The following components compose the SP/CP Subsystem:

- Service Registry: it is dedicated to store metadata and references for the Media Services/Contents available to the End-Users and to allow discovery and retrieval. In order to make the Media Services/Contents available for retrieval, they need to be registered into the Service Registry. In order to enable End-Users to discover a service, a query shall be sent by the End-User via his/her Home-Box (Home-Box Services module) to the Service Registry module through the interface, named SHr. The query is likely to contain contextual End-User information (e.g., User preferences and Terminal capabilities);
- Service Composition: it is responsible for provisioning Media Services into the CP subsystem in three different ways: (1) imported from the existing applications such as the existing IPTV services provided by the Legacy IPTV Platform, (2) created by SP and defined using the service composition module such as the Composite Media Service, (3) created by the End-Users with their UGC (User Generated Content), registered and published into the Service Registry in order to be accessible by other End-Users;
- Service Management: this component assures the efficient cooperation between the SP and NP, enabling the capabilities of creating Virtual Content-Aware Networks. It also handles all the services subscription mechanisms, the administration and control of the Content Media Servers, the Service Monitoring features and all the lifecycle management of Media Services (creation, planning, publishing, delivery, etc.);
- Service Delivery: this module enforces the delivery of Media Contents to the End-Users, through streaming protocols, such as HTTP (DASH-based) or RTP;
- Content Server Adjustment (CSA): this module’s role (in cooperation with the management module) is to trigger Content Server adjustment actions which jointly provide improvement to the system load conditions in different timescales and optimization of the whole adaptation process.

The SP/CP Subsystem brings innovation in the field of service management, by proposing a **“Cooperative Service Management for Emerging Media Applications”**, where the innovations points stand within the dual adaptation mechanism (content adaptation & server adjustment) and the flexible mechanism for cooperation between the Service Provider and Network Provider (CAN Provider) in order to provide the **Content-Awareness** feature at the Network level. This contribution is not detailed in this document but a short summary is provided in Annex A, including links for deeper information.

2.4.2.3 The Home-Box Subsystem

The Home-Box Subsystem aims at allowing Service Providers to deliver enriched networked Media Services and Contents that can be efficiently exploited by End-Users. It acts as the logical interconnection of deployed Home-Boxes, so it will represent, for the Service Providers and End-Users, a virtualized layer capable of advanced ways of Service/Content provisioning along with **Context+Network-Awareness** capabilities and providing the inputs for achieving **Content-Awareness** at the Network level.

The Home-Box, as the main element of the HB layer, can be seen as the evolution of today's Home Gateways (HGs), with advanced building blocks such as User Management, Service Management and Middleware, for deploying innovative features in terms of Monitoring, Adaptation, and Delivery (Caching).

As a whole, the virtual Home-Box layer enables different modes of media distribution and delivery, including traditional Client/Server streaming mode (HTTP or RTP streaming), multicast mode for multicast-enabled domains, and peer-to-peer mode for domains without multicast support. The Home-Box is enhanced with the content caching and forwarding (based on the video popularity, for example) and the capability of receiving pushed content from the SP, so that the Home-Box virtual overlay can assist the current CDNs (**HB-assisted CDN**).

In few words, via the HB layer, the HB takes into account user dynamic context information to support the user **context-aware applications, enhanced Quality of Experience (QoE) and intelligent adaptation**. With that purpose in mind, a multi-layer monitoring subsystem has been designed and specified in the HB entity. Through it, it becomes possible to collect context information from the End-User terminals (EUTs) (i.e., Context-Awareness feature), provide it to the HB, and also retrieve network and HB distances from the CAN for best end-point selection during the service consumption phase (i.e., **Network-Awareness feature**).

Since one of the main objectives is to distribute content in a very efficient way, so that the network resources may be optimized (avoiding redundancy and reducing the amount of requests to be processed by SP/CP servers), the appropriate approach is to create a distributed infrastructure, whose leaves are the HBs. This way, media content distribution can benefit from: **(1) the employment of caching in the HB nodes and, (2) the construction of a mesh network exploiting P2P technology.**

Another advantage of the approach above is related to QoS/QoE. Taking into account the amount of different terminals that might exist in the deployed configuration, each one with their own characteristics and capabilities, it is unlikely that a single SP/CP could eventually produce all the necessary service/content formats to serve them, so it could be suitable for every single terminal and, at the same time, providing the best possible QoS/QoE. For that reason, our approach to solve the issue is via the introduction of **a two-layer media content adaptation, taking both advantages of the NP infrastructure adaptation at the network level and also at the HB virtual layer.**The adaptation architecture must be, therefore, distributed compared to other possible solutions where the CP, or even the SP, produces all the required content formats by himself, in a centralized (or less distributed) approach, which makes our adaptation architecture scalable and very flexible.

Existing HG specifications (1) do not support the challenges discussed in chapter 2, such as the context/network-aware service management and media adaptation, the QoE/Distance-aware end-to-end/cross-layer monitoring; (2) are not able to take advantage of the content-aware networking and virtualization. These two aspects are major drivers towards service assurance and improving the End-Users' experience. In this context, the Home-Box (HB) entity we propose, aims to constitute a next generation HG providing means for **dynamic X-Aware service management/delivery, monitoring and adaptation support**. In order to support all the aforementioned features, the HB consists of:

- ***HB User Management*** functional block for session instantiations, according to the User Profile and Context. It includes a session management function, essential in context of multiple users and multiple service instances and a centralised User Profile manager (receiving input from user's terminal management modules), so that the HB has a unified view of all registered users/terminals:
  - ***Related innovation: "Context–Aware Framework for Media Services Provisioning".*** The Home-Box includes functionalities to consider EU's varying context information (context storing and management, monitoring), for service **personalization** and **adaptation** support, including Quality of Experience (QoE) evaluation. This contribution is presented in Chapter 3;
- ***HB Service Management*** functional block using the middleware to enable HB services for the End-Users and providing the following functionalities: Delivery, Discovery, Publishing, Gateway, and Management:
  - ***Related innovation: "Home-Box Assisted Content Delivery Network".*** The Home-Box includes media server functionalities for directly serving content, with or without any mediation entity (e.g., a media aggregator platform). Flexible communication modes (unicast (DASH or RTP), P2P modes) are present allowing to act as a powerful media server entity for SPs/CPs or for End-Users' service consumption and delivery. A specific caching algorithm is used and replication may be performed online as well as offline. The delivery phase is preceded with the discovery phase where innovative features are also proposed based on **Network- & Context-Awareness** (based on the **Context-Aware Framework** pre-cited). This contribution is presented in Chapter 4;
- ***HB Middleware*** functional block implementing the core operations of the HB entity and layer, such as Monitoring and Adaptation along with their related innovative features (as presented in Annex C).

In conclusion, the HB extends the existing HG specifications, by enhancing the previously identified HG functional blocks with new innovative functionalities, in order to progress in the aforementioned Research Challenges.

2.4.2.4 The CAN & Network Infrastructure Subsystem

The CAN & Network Subsystemoffers to the SP Virtual Content Aware Network services with different levels of QoS guarantees based on virtual networks in the Data Plane (mono or multi-domain, in unicast or multicast mode) and constructed through Service Level Agreements (SLA) contracts concluded in the Management Plane. It also provides network information to the Home-Box so that they can enable the **Network-Awareness** feature. And most of all, it executes **Content-Awareness** features inside the MANE network elements. The components permitting these features are:

- The Multi-domain CAN Manager: it instantiates the Virtual Content-Aware Network through multiple domains and multiple providers;
- The MANE (Media-Aware Network Element): it is the building block of the VCAN layer, which can be seen as the evolution of today’s edge routers, with advanced functionalities of Content-Awareness, including forwarding, monitoring and adaptation inside the network;
- The Intra-NRM (Intra Network Resource Manager): it is the component having the ultimate authority for configuring/controlling the network resources;
- The Core Network Nodes: they represent common network routers, implementing Diffserv/MPLS for achieving QoS requests.

The CAN & Network Subsystem brings innovation in two planes:

- In the Data Plane by proposing **Content-Awareness** with a network elements (incl. the Media-Aware Network Element (MANE)), functions and algorithms (for unicast and multicast flows);
- In the Control Plane by proposing a **Management System for Virtual Multi-Domain Multi-Provider Content-Aware Networks (for unicast and multicast)**.

These contributions are not detailed in this document but a short summary is provided in Annex C, with details on related innovations: in the Data Plane (for **Content-Awareness** mechanisms) and in the Control Plane (for the **multi-provider, multi-domain management system**).

2.4.2.5 The Distributed Adaptation System

The Distributed Adaptation System uses a cross-layer distributed in-network adaptation solution in order to support (1) improvement of QoE; (2) management of network traffic; (3) universal access to the content for multiple devices.

As shown in **Figure 4**, adaptation is deployed at several points in the system: the CP, the HB, the MANE, and the CAN Manager. The high-level Adaptation Framework (AF) architecture comprises two essential components: the Adaptation Decision-Taking Framework (ADTF) and the Processing Engine/Adaptation Engine (PE/AE). The ADTF is based on the approach of Digital Item Adaptation (DIA) introduced in MPEG-21 [27]. Based on that adaptation decision, the PE/AE performs the actual adaptation.

The Processing Engine (PE) is located at the HB and at the CP and performs encoding (to SVC), decoding (from SVC), and adaptation (including transcoding from SVC to X and from X to SVC). The Adaptation Engine (AE) is located at the MANEs and is the counterpart to the PE. The AE only performs SVC adaptation, by dropping layers.

The Distributed Adaptation System brings innovation (1) in the decision process and (2) in the execution process. We propose a **"Distributed Framework for Edge and In-Network Media Adaptation"**, enabling Universal Media Access to End-Users and optimization of resources to Content Providers and Network Providers. This contribution is detailed in Chapter 5.

2.4.3 Comparison with existing solutions

Our Future Internet architecture towards a Media Ecosystem (NetME) can be seen as a solution deriving from the FMIA-TT reference model. Even though, the mapping between NetME and FMIA-TT reference model strata is not exactly one-to-one since NetME does not have in its scope all FMIA-TT functions, the correlations are the ones presented in **Table 1**.

Table 1: FMIA-TT and NetMEArchitectures Mapping

FMIA-TT Architecture	*NetME Architecture*
SNI	Network Environment+Service Environment
DSCAO	CAN Layer
IO	HB Layer
AL	User Environment + Service Environment

Therefore, a parallel can easily be setup between the two approaches. Indeed, we have collaborated within FMIA-TT, bringing new aspects, through the European Project ALICANTE.

In addition, **Table 2** presents a comparaisson between our solution and the state-of-the art solutions provided in Section 2.3.2. More details including other technical and business points of view can be found in [3].

Table2: Comparison between NetME and other Architectures Proposals

Main Features	MESCAL	ENTHRONE	AGAVE	4WARD	PSIRP/ PURSUIT	COMET/ CURLING	OCEAN	NetME
Media (including real-time) services oriented	No	Yes	Yes	No	No	Yes	Yes	**Yes**
Full high-level services management	No	Yes	No	Yes	Yes	Yes	No	**Yes**
Multi-domain	Yes	Yes	Yes	Yes	Yes	No	No	**Yes**
Network Virtualisation	No	No	Yes	Yes	No	Partial	No	**Yes**
CAN/NAA concepts	No	No	No	No	No	Yes	Yes	**Yes**
Network Resource Management	Yes	Yes	Yes	Yes	No	Yes	Yes	**Yes**
P2P capabilities	No	No	No	No	No	No	No	**Yes (P2P-Next)**
Home-Box concept and HB virtual layer	No	No	No	NA	NA	NA	No	**Yes**

2.4.4 Summary of contributions

The contributions performed in relation with the proposed architecture are summarized in **Figure 5**. This figure maps the research contributions and the "Research Challenges" they address into the different environments and related components of the NetME architecture.

In the following two chapters, we will mainly focus on presenting two key contributions in relation with the Home-Box layer (the others can be found in Annexes and related research papers/deliverables), namely:

1. A Context–Aware Framework for Media Services Provisioning (Chapter 3);
2. A Home-Box Assisted Content Delivery Network System (Chapter 4).

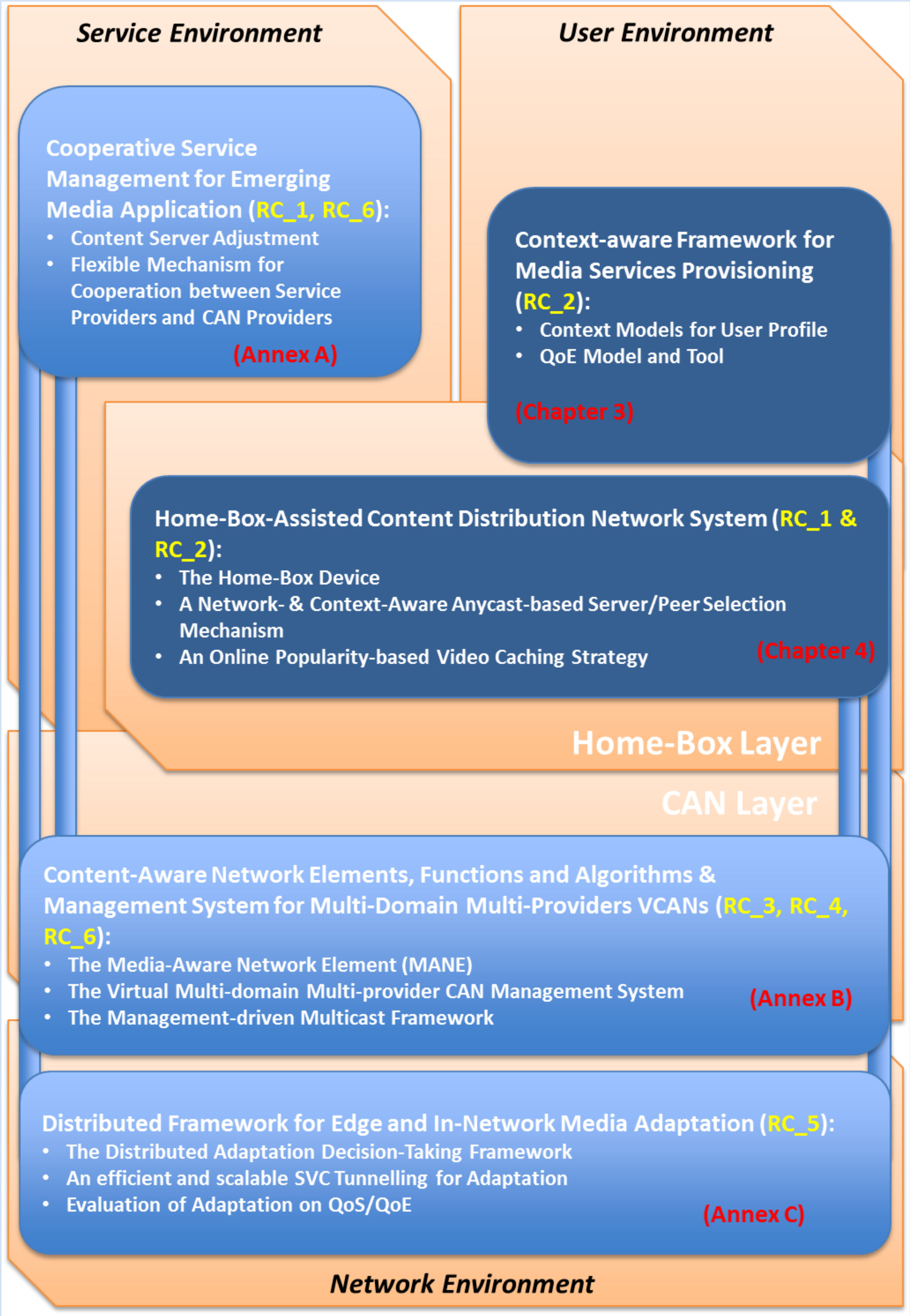


Figure 5: Summary of Contributions

Bibliography

[1] A. Galis et. al., "Management and Service-aware Networking Architectures (MANA) for Future Internet Position Paper: System Functions, Capabilities and Requirements", http://www.future-internet.eu/home/future-internet-assembly/prague-may-2009

[2] FIArch Group: Fundamental Limitations of Current Internet and the path to Future Internet (March 2011), http://ec.europa.eu/information_society/activities/foi/docs/current_internet_limitations_v9.pdf

[3] G. Xilouris, D. Negru (eds.) et al., ICT **ALICANTE**, Deliverable D2.1: "ALICANTE Overall System and Components Definition and Specification", September 2011, http://www.ict-alicante.eu/validation/download/work-package/alicante_d2.1_final.pdf

[4] FIArch Group: Fundamental Limitations of Current Internet and the path to Future Internet (March 2011), http://ec.europa.eu/information_society/activities/foi/docs/current_internet_limitations_v9.pdf

[5] Future Media Internet Architecture Think Tank (FMIA-TT), "Future Media Internet Architecture Reference Model", 1. March 2011, http://www.future-internet.eu/uploads/media/FMIA_Reference_Architecture.pdf

[6] M. Baldauf, S. Dustdar, and F. Rosenberg, "A survey on context-aware systems", International Journal of Ad Hoc and Ubiquitous Computing, 2(4):263–277, 2007

[7] G. Chen and D. Kotz, "A survey of context-aware mobile computing research", Technical Report, Dartmouth College, Hanover, NH, 3755, 2000

[8] S. Mostefaoui and B. Hirsbrunner "Towards a context-based service composition framework" Proc.1st International Conference on Web Services (ICWS 03), 2003

[9] Y. Yang, F. Mahon, M. Williams, and T. Pfeifer, "Context-aware dynamic personalised service re-composition in a pervasive service environment", Ubiquitous Intelligence and Computing, pages 724–735, 2006

[10] ICT/FP7 C-CAST, http://www.ict-ccast.eu

[11] S. Jin, A. Bestavros, and A. Iyengar, "Accelerating internet streaming media delivery using network-aware partial caching", Proc. ICDCS '02 Proceedings of the 22 nd International Conference on Distributed Computing Systems (ICDCS'02), 2002

[12] B. Tierney, J. Lee, B. Crowley, M. Holding, J. Hylton, and F. Drake Jr., "A network-aware distributed storage cache for data intensive environments", Proc. Eighth International Symposium on High Performance Distributed Computing, 1999

[13] IETF ALTO Working Group. http://tools.ietf.org/wg/alto/

[14] V. Aggarwal, A. Feldmann, and C. Scheideler, "Can ISPs and p2p users cooperate for improved performance?", ACM SIGCOMM Computer Communication Review, 37(3):29–40, 2007

[15] D. Saucez, B. Donnet, and O. Bonaventure, "Implementation and preliminary evaluation of an ISP-driven informed path selection", Proc. 2007 ACM CoNEXT conference, page 45. ACM, 2007

[16] H. Yin, X. Liu, T. Zhan, V. Sekar, F. Qiu, C. Lin, H. Zhang, and B. Li. Livesky, "Enhancing CDN with p2p", ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), 6(3):1–19, 2010

[17] ICT/FP7 ADAMANTIUM, http://www.ict-adamantium.eu

[18] 3GPP TS 23.228 "IP Multimedia Subsystem (IMS); Stage 2.", 3rd Generation Partnership Project (3GPP), Sept. 2010

[19] A. Dey, "Understanding and Using Context", ACM Personal and Ubiquitous Computing V. 5, N. 1, Springer-Verlag, Atlanta, GA, USA, 2001, pp: 4-7

[20] M. Rockl, K. Frank, P.G. Hermann, M. Vera, "Knowledge Representation and Inference in Context-Aware Computing Environments", Proc. Second International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, Valencia, Spain, 2008

[21] The CC/PP Working Group, http://www.w3.org/Mobile/CCPP/

[22] 3GPP, "Generic User Profile (GUP)", 3GPP TS 22 240 v8.0.0, 2008-12

[23] 3GPP, "User Data Convergence - Data Model", completed as Release 10, 2011

[24] ITU Draft Rec. Y.2012- "Functional Requirement and architecture of the NGN", release 1

[25] 3GPP, "Telecommunication management; Study of Common Profile Storage (CPS) Framework of User Data for network services and management", 3GPP TR 32.808, Release 8, 2007

[26] IEEE Draft Standard "Next Generation Service Overlay Network", IEEE P1903/D1, April 2011

[27] ISO/IEC 21000, "Information Technology – Multimedia Framework (MPEG-21)"

[28] R. Kernchen , David Bonnefoy , A. Battestini , B. Mrohs , M. Wagner , M. Klemettinen, "Context-Awareness in MobiLife", Proc. IST Summit, 2006

[29] The FP7 PERSIST Project: "Personal self-improving smart spaces". http://www.ict-persist.eu/

[30] COAST, http://www.coast-fp7.eu/

[31] C. Perkins, P. Calhoun, J. Bharatia, "IP Mobility Support for IPv4", RFC 4721, January 2007

[32] "Media independent handover (MIH)", IEEE 802.21. http://www.ieee802.org/21/

[33] 3GPP, "IP Multimedia (IM) Core Network (CN) subsystem IP Multimedia Subsystem (IMS) Service Continuity", 3GPP TS 24.237 Release 10.3.0, 2011

[34] IST-FP6 Mobilife project, http://research.nokia.com/page/239

[35] FP7 PERIMETER: "User-centric paradigm for seamless mobility in future Internet". http://www.ict-perimeter.eu/

[36] U. Engelke, H.-J Zepernick, "Perceptual-based quality metrics for image and video services: A survey", Proc. 3rd EuroNGI Conference on Next Generation Internet Networks, Trondheim, Norway, 2007

[37] S. Mohamed, G. Rubino, "A study of real-time packet video quality using random neural networks", IEEE Transactions on Circuits and Systems for Video Technology, vol. 12 , pp. 1071 – 1083, 2002

[38] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, S. Moon "Analyzing the Video Popularity haracteristics of Large-Scale User Generated Content Systems", IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 17, NO. 5, OCTOBER 2009

[39] COMET, http://www.comet-project.org/

[40] B. Subbiah, Z. Uzmi, "Content aware networking in the Internet: issues and challenges", Proc. IEEE Int. Conf on Communications, 2001, vol.4, no., pp.1310-1315, doi:10.1109/ICC.2001.936912

[41] A. Carzaniga, A. L. Wolf, "Forwarding in a Content-Based Network", Proc. SIGCOMM 03, www.inf.usi.ch/carzaniga/papers/cw_sigcomm03.pdf

[42] MESCAL D1.2: "Initial Specification of Protocols and Algorithms for Inter-domain SLS Management and Traffic Engineering for QoS-based IP Service Delivery and their Test Requirements", January 2004, www.mescal.org

[43] T.Ahmed, A. Asgari, A.Mehaoua, E. Borcoci, L.B Équille, K. Georgios "End-to-end quality of service provisioning through an integrated management system for multimedia content delivery" Computer Communications, Special Issue: Emerging Middleware for Next Generation Networks, Volume 30, Issue 3, 2 February 2007, Pages 638-651

[44] M. Boucadair, P. Lévi, D. Griffin, N. Wang, M. Howarth, and G. Pavlou, E. Mykoniati and P. Georgatsos, B. Quoitin, J. Rodríguez Sánchez, M. L. García-Osma, "A Framework for End-to-End Service Differentiation: Network Planes and Parallel Internets", IEEE Communication Magazine, Sept. 2007, pp. 134-143

[45] M.Boucadair, (ed) al., AGAVE Public Deliverable "D1.1: Parallel Internets Framework", September 2006, www.ist-agave.org

[46] 4WARD, "A clean-slate approach for Future Internet", http://www.4ward-project.eu/

[47] F.J. Ramón Salguero, "COntent Mediator architecture for content-aware nETworks", 2010, http://www.comet-project.org/

[48] Gerardo García de Blas, (Ed) et al., Public Deliverable D2.1, "Business Models and System Requirements for the COMET System", 2011, http://www.comet-project.org/

[49] Publish-Subscribe Internet Routing Paradigm, http://www.psirp.org/

[50] M. Ain ed. et. al., Public Deliverable D2.2, "Conceptual Architecture of PSIRP" http://www.psirp.org/

[51] Nikos Fotiou, et. al., "Developing Information Networking Further: From PSIRP to PURSUIT"

[52] W. K.Chai, Ning Wang, I.Psaras G.Pavlou, C.Wang, G. García de Blas, F.J. Ramon Salguero, Lei Liang, S.Spirou, Andrzej Beben, E. Hadjioannou, "CURLING: Content-Ubiquitous Resolution and Delivery Infrastructure for Next-Generation Services" , IEEE Communications Magazine, March 2011, pp.112-120

[53] J. Choi, J.Han, E. Cho, T. Kwon, and Y. Choi, "A Survey on Content-Oriented Networking for Efficient Content Delivery", IEEE Communications Magazine • March 2011, pp. 121-127

[54] Howarth, M.P. et al., "Provisioning for Interdomain Quality of Service: the MESCAL Approach", IEEE Communications Magazine, June 2005, pp. 129-137

[55] ENTHRONE, End-to-End QoS through Integrated Management of Content, Networks and Terminals, FP6 project, URL: http://www.ist-enthrone.org/

[56] T. Anderson et. al., "Overcoming the Internet Impasse through Virtualization", Computer, vol. 38, no. 4, Apr. 2005, pp. 34–41

[57] J. Turner and D. Taylor, "Diversifying the Internet," Proc. GLOBECOM '05, vol. 2, St. Louis, USA, Nov./Dec. 2005

[58] N. M. Chowdhury and R. Boutaba, "Network Virtualization: State of the Art and Research Challenges", IEEE Communications Magazine, vol. 47, no. 7, Jul. 2009, pp. 20-26

[59] F. Verdi, Maurıcio F. Magalhaes, "Using Virtualization to Provide Interdomain QoS-enabled Routing", Journal of Networks, April 2007, pp. 23-32

[60] Open ContEnt Aware Networks (OCEAN), http://www.ict-ocean.eu/

[61] A. Menezes, P. van Oorschot and S. Vanstone, "Handbook of Applied Cryptography",

CRC Press, Oct. 1996

[62] W. Stallings, L. Brown, "Computer Security. Principles and Practice", Ed. Prentice Hall, ISBN 0-13-600424-5, 2008

[63] J. Mirkovic, S. Dietrich, D. Dittrich, P. Reiher, "Internet Denial of Service: Attack and Defense Mechanisms", Ed. Prentice Hall, ISBN 0-13-147573-8, Dec. 2004

[64] S.Kent, K.Seo, "Security Architecture for Internet Protocol", RFC 4301, December 2005

[65] T. Wiegand, L. Noblet and F. Rovati, "Scalable Video Coding for IPTV Services", IEEE Transactions on Broadcasting, 2009, 55, 527 -538

[66] S. Wenger et al., "RTP Payload Format for Scalable Video Coding", IETF Internet Draft, http://tools.ietf.org/html/draft-ietf-avt-rtp-svc-27

[67] H. Schwarz, D. Marpe and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard" IEEE Transactions on Circuits and Systems for Video Technology, 2007, 17, 1103 -1120

[68] S. Dobson et al. "A survey of autonomic communications", ACM Transactions on Autonomous and Adaptive Systems (TAAS), vol. 1, Issue 2, pp. 223 – 259, 2006

[69] E. Gelenbe, "Steps toward self-aware networks," ACM Communications, Vol.52, No.7 (Jul. 2009), pp.66-75

[70] "Evolving the Internet" Presentation to the OECD, March 2006, http://www.cs.ucl.ac.uk/staff/m.handley/slides/

[71] M. Handley, "Re-thinking the control architecture of the internet". Keynote talk, ACM ReArch Workshop 2009 (co-located with ACM CoNext 2009), December 2009, http://www.cs.ucl.ac.uk/staff/m.handley/slides/rearch-ho.pdf

[72] Theodore Zahariadis, Dimitri Papadimitriou, Hannes Tschofenig, Stephan Haller, Petros Daras, George D. Stamoulis, and Manfred Hauswirth. "Towards a future internet architecture" The future internet, John Domingue, Alex Galis, Anastasius Gavras, Theodore Zahariadis, and Dave Lambert (Eds.). Springer-Verlag, Berlin, Heidelberg 7-18

[73] Wei Koong Chai, Ning Wang, I. Psaras, G. Pavlou, Chaojiong Wang, G. G. de Blas, F. J. Ramon-Salguero, Lei Liang, S. Spirou, A. Beben, E. Hadjioannou, "Curling: Content-ubiquitous resolution and delivery infrastructure for next-generation services," Communications Magazine, IEEE, vol.49, no.3, pp.112-120, March 2011

[74] N. Ramzan, E. Quacchio, T. Zgaljic, S. Asioli, L. Celetto, E. Izquierdo, F. Rovati, "Peer-to-peer streaming of scalable video in future Internet applications," Communications Magazine, IEEE, vol.49, no.3, pp.128-135, March 2011

[75] H. Koumaras, D. Négru et al., "Media Ecosystems: A Novel Approach for Content-Awareness in Future Networks", Future Internet: Achievements and Promising Technology, Springer Verlag, pp.369-380. May 2011

[76] Cheok, Adrian David, "Art and Technology of Entertainment Computing and Communication", ISBN 978-1-84996-136-3. Springer-Verlag London Limited, 2010, pp. 59

[77] M. Rockl, K. Frank, P.G. Hermann, M. Vera, "Knowledge Representation and Inference in Context-Aware Computing Environments", Proc. Second International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, Valencia, Spain, 2008

[78] "P2P-Next, Next Generation Peer-to-Peer Content Delivery Platform", http://www.p2p-next.org

Chapter 3

Context–Aware Framework for Media Services Provisioning

Meeting End-Users needs, interests and preferences is becoming a key competitive factor for Content and Service Providers. Despite the progress since the last decade, providing End-Users with a strong and compelling media experience remains a challenging issue. We tackle this need by proposing a Context-Aware Framework, capable of handling the richness of the computing context towards personalized Media Services and better End-Users experience**.**

3.1 Introduction

With the compelling proliferation of smart devices, the rapid growth in wireless networking technologies and the diversification of Media Services, users require more ubiquity and personalization in the service access [1].

One key technology enabling ubiquitous and seamless access to services is the Context-Awareness. This promising technology enables the tracking of context information from different sources in the network, the modeling of content information to facilitate its processing by software entities, and the understanding of this information to take adaptation decisions on the applications' behavior, in order to provide End-User with services, contents and resources relevant to their current situation, without explicit intervention from them.

Indeed, Media Services are known to be very sensitive to diverse context information related to different context entities. Constraints induced by user, device, network and environment context have a strong impact on the efficiency of the system and the quality perceived by the users. Therefore, the design of an efficient Context-Aware framework should address several issues:

- *What pertinent context information to consider?* Depending on the domain of interest;
- *What model to use to formally describe context information?* Define the trade-off between complexity and expressivity;
- *How and where the context information should be managed?* The context management should be defined with keeping in mind the scalability, extensibility and responsiveness features; and it should be decided if it should be managed at the User Environment, the Service Environment or within an intermediate node to be defined.

The following details the contribution towards Context-Awareness, in conjunction with the Home-Box layer by presenting, **first, the proposed context models** (using XML and ontologies) and, **second, the QoE model** (predominant information of the context) used to evaluate the End-Users' ongoing Media Experience.

3.2 The Context Models

3.2.1 Related Works

Today, ubiquitous computing [2] becomes reality. The idea of ubiquitous computing is to put the software components and the networked computing devices that implement the services into the background towards focusing on providing End-Users with high quality service anytime, anywhere and through any device or networking technology. Indeed, from the End-User point of view, only the service value counts, not the networked device or software components that implement it.

One key aspect enabling ubiquitous services is Context-Awareness. The promise of Context-Awareness is to provide computing frameworks that, thanks to different sensing technologies, track the user context, understand enough on it and adapt applications behaviour accordingly to provide the End-User with contents, resources and services relevant to his current situation without explicit intervention from him. Context could be a user location, preferences and activity, device capabilities, network conditions, environment information such as time, light intensity, motion, sound noise level, etc.

The motivation for Context-Awareness is gaining importance with the vision of future media Internet that is foreseen to fully handle a wide range of Media Services (VoD, IPTV, VoIP, gaming applications, and many others to come). Indeed, Media Services are known to be very sensitive to diverse context information related to different context entities. Constraints induced by user, device, network and environment context have a strong impact on the efficiency and appropriateness of their provisioning. Media Services, more than others, need then to benefit from a context-aware framework that enables dynamic and automatic service personalization and content adaptation toward a better End-User experience. However, with the explosive demand experienced by Internet Media Services and the wide context diversification inducing huge processing capabilities needed in such a framework, it is obvious that scalability is one of the most important requirements of the design of such a framework.

Context modelling is a fundamental field in Context-Awareness concept. Indeed, to be processed by computational entities, context information should be formally described in a contextual model. Strang and Linnhoff-Popien classified context models based on the data structure used to represent and exchange context between the system entities. They identified six models: key-value models, markup scheme models, graphical models, object-oriented models, logic-based models and ontology-based models.

The *key-value models* are undoubtedly the models with the simplest data structure. The context information is represented as key-value pairs (name of the context information and its actual value). The simplicity of the key value data structure facilitates the management of context information. Unfortunately, its lack of expressiveness prohibits any deduction from the considered context information and its flat structure does not support the relationship definition among parameters. The absence of data schema and meta-information on the considered context makes this type of models very difficult to reuse. Indeed, the key-value models are usually tightly coupled with the systems for which they have been built for.

The *markup scheme models* are characterized by a hierarchical data structure. The context information is organized into elements identified by their tags, which are associated with attributes and contents. Recursively, an element can itself contain other elements. These models are often used in the standard for user profiling. Markup scheme-based model are the mostly adopted models to represent context. There exist many standards and tools (parsers, validation tools, etc.) that are based on it, especially with the generalization of the use of XML within Web services tools and standards. In addition, the herarchical structure on which this model is built fits well the decomposition nature of content information. However, existing standards are always targeting a specific type of applications and context domain. For example, MPEG is specific to user and content, CC/PP is user and device orientied, etc. None has reached a full generic solution capable of meeting all requirements for a generic User Profile. Another limit of these models is that they only model the syntax of context information and lack for semantic. There is no way to represent meta-information or to model the relationship that may exist between.

Two solutions for *graphical model* are highlighted:

- A generic and well known modelling instrument that is also appropriate to model context is the Unified Modelling Language (UML). The context entities and their processing are represented in the UML diagrams (class diagram, use case diagram, sequence diagram, etc.). An example of UML-based graphical model is the air traffic control presented in [3];
- Another graphical model for modelling context is the one introduced in [4], in which Henricksen et al. extend the Object Role Modelling (ORM) [5] to allow context facts types to be categorized according to their persistence and source. In ORM, the basic modelling concept is the *fact*, and the modelling of a domain using ORM involves identifying appropriate fact types and the roles that entity types play. Facts are classified as either static or dynamic. The latter ones are, in their turn, classified as profiled, sensed or derived.

The strengths of graphical models are their efficiency in representing the structure of context information. These models are intuitive and easy to integrate to the UML model of the rest of the system. Some code can also be derived from the context diagrams. However, since the graphical models are commonly used for human structuring purpose, they present a low level of formalism.

The *object-oriented models* approach encapsulates the representation and process details of context entities (such as location, identity, etc.) in context objects. The latter are accessed via well-known interfaces. The advantage of using such an approach in context modeling is to benefit from the full power of the object oriented approach (e.g., encapsulation, inheritance, reusability).

In *logic-based models*, context is defined using facts, expressions and rules. The high level formality of this type of models allows high-level reasoning or inference. Context information can then be added to, modified and deleted from the logic-based system in terms of facts or derived from the system rules.

The *ontology-based models* represent the context based on ontologies [6]. Originally, the ontology was defined, in philosophy, as the study of the nature of *being, existence* or *reality*, as well as the basic categories of being and their relations. Ontologies are used in Computer Science for formal system representation using concepts, attributes and relations. By providing an explicit conceptualisation giving a description of data structure and semantics, ontologies are perceived as promising tools towards adequate description and representation of context data. Their relation with the semantic Web also constitutes an important factor in the fact that different languages have been defined for ontologies, e.g., Ontolingua [7], LOOM and Ontology Web Language (OWL) [8].

3.2.2 Evaluation, Choice and Architecture of the Context Models

Evaluation

An evaluation of the six surveyed models is presented in [9]. The evaluation is based on six requirements that are defined as fundamental in ubiquitous computing:

- Distributed composition (dc): since computing systems are usually distributed, this feature is important due to the lack of a central instance being responsible for the creation, deployment and maintenance of data and services, in particular context descriptions;
- Partial validation (pv): as a result of the distributed composition requirement, the partial validation of contextual knowledge is particularly important;
- Richness and quality of information (qua): a context model should inherently support some quality and richness indicators (e.g., uncertainty, accuracy, etc.) for context information;

- Incompleteness and ambiguity (inc): being usually incomplete and/or ambiguous, this limit should be covered by the context model;
- Level of formality (for): it is important that the different computational entities composing the system have a shared understanding of a domain vocabulary;
- Applicability to existing environments (app): from the implementation perspective, it is important that a context model is applicable within the existing infrastructure.

The results of the study are summarized in **Table 3**. The conclusion is that ontology-based models are the ones that suit best the requirements of context modeling in ubiquitous computing, especially the formality, distributed composition and partial validation. Ontology-based models have also received much attention in the last years due to their linkage with the semantic Web.

Table 3: Context Models' Appropriateness Indication

Approach/requirements	dc	pv	qua	inc	For	app
Key-Value Model	-	-	-	-	-	+
Markup scheme Model	+	++	-	-	+	++
Graphical Model	-	-	+	-	+	+
Object-oriented Model	++	+	+	+	+	+
Logic-based Model	++	-	-	-	++	-
Ontology-based Model	++	++	+	+	++	+

<u>Choice</u>

Based on above conclusions, two model solutions have been chosen for representing context, the markup schemes-based model and ontology-based model. The first one is the mostly used in benefits from the mature XML-based languages, technologies and tools. The second one is more expressive and formal enabling advanced functionalities such as reasoning. However, this comes with more complexity which results on less responsiveness. The latter limits make this type of models difficult to apply to multimedia frameworks that are known to be time-constrained. We have then proposed two context-aware frameworks based on these two models. In the first contribution, the context model is based on XML and its dialects and the context management relies on XML related tools. Based then on markup scheme models, this contribution aims to provide a context-aware multimedia framework with high performance. In the second contribution, the context is modeled using ontology and rules for a more expressiveness and formality. The aim was to prove the feasibility of such models in the field of large-scale Media Services. Since the context model support reasoning, this function is incorporated in the context management to ensure context consistency checking, high-level context inference and context-aware decisions triggering.

Architecture

As afore mentioned in chapter 2, the proposed NetME architecture aims to provide the Context-Awareness and the associated media content personalization and adaptation features on top of the Home-Box entity so as to enable ubiquitous access to context-adaptive Media Services without endanger the scalability and performance features (**Figure 6**).

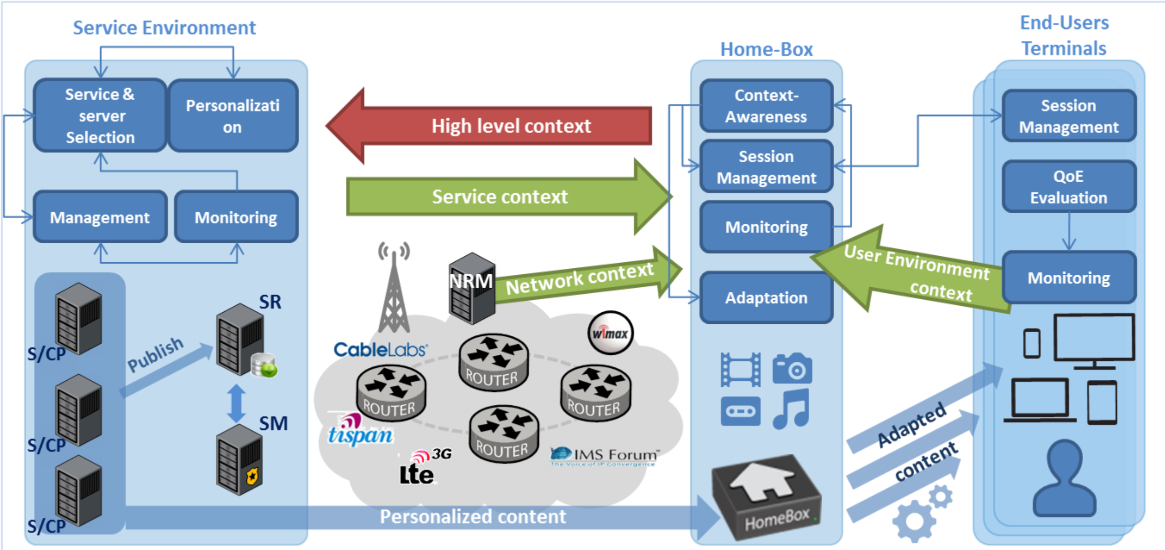


Figure 6: Functional Architecture regarding Context-Awareness

The HB is enhanced with the following functionalities (depicted in **Figure 6**):

- ***Context management***: the HB layer is responsible for the Context-Awareness feature. Context information acquired from different environments (User, Service and Network Environments) is then abstracted to be integrated in a formal markup or ontology-based model permitting it to be shared between different users, devices and services. Situational context is also deduced or inferred depending on the used model to dynamically supply services with accurate high-level context information, when needed. The distribution of context management among HBs has several advantages as follows:
    - It relieves the resource-constrained devices for the task of context reasoning;
    - Scalability can be achieved by distributing the context management features among Home-Boxes;
    - The Home-Box is the central element in home network, so besides managing context for Internet services, it can also manage the home appliances and services;
    - Being an intermediate equipment that can be easily reached from any of the User, Service and Network Environments (identified in the architecture) and implementing itself the major context consumers processes (content personalization, adaptation and service and session management), the HB

constitutes an ideal place to which the monitoring information can converge while minimizing communication overhead;

- ***Service personalization and content adaptation**:* to overcome the high context heterogeneity, adaptation seems to be an effective solution. Multimedia contents should be adapted to always meet the computing device capabilities, the access network conditions or the supported format by the displaying applications ensuring thus to satisfy service quality. Some adaptation actions, such as the selection of the right content format when it is available, can be achieved at the Service/Content Provider (SP/CP) side. However, adaptation actions such as resizing and transcoding can't be performed neither in a centralized way (at the CP side) nor at the End-User terminal side due to the huge processing capabilities that they require. Such actions are performed at the HB entity that will act as an intermediate node for its associated User Environment. Another trend goes in the direction of Media Services personalization towards providing End-Users with appropriate and enriched contents that fit their preferences and interests. In the proposed framework, this feature is achieved in collaboration between the HB, that sets the parameters of the discovery request based on the User Profile, and the Service Registry (SR), that performs the matching of the request with the service' descriptions previously published by Service Providers (SPs). The decisions taken within these processes could be either invoked (markup schemes) or directly derived from context by means of reasoning or a result of some algorithms whose execution will be triggered by events sent by the context-aware feature (ontology-based schemes) ;
- ***Session Management**:* since the objective is to abstract as much as possible the dynamicity of the context-based adaptive behaviour of the system to the Service Environment for more scalability, the HB is acting as a proxy of its related devices for all the signalling process. The User Environment context-driven adaptive management actions are hence performed locally. An example of such actions can be the achievement of multiple device deployment by enabling a seamless handover of the session from one device to another, when the user moves.

To ensure high service quality and keep the system under control, software ***monitoring*** runs in all the system nodes to collect computational and connectivity context. The reporting can be periodic or on-demand according to the system needs. The software monitoring is also coupled with a subjective quality evaluator tool [10] at the End-User device side that dynamically evaluates the quality perceived by the user.

The following sub-sections present the two proposed frameworks able to achieve Context-Awareness.

3.2.3 Markup scheme-based context framework

The pertinent context information that constitutes the overall markup scheme-based model is the User Profile. Dey [11] defines context as any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves. Based on this definition, four entities have been identified as relevant to the User Profile design, namely the User, Device, Service and Context. Each entity is represented as a sub-profile. The first three sub-profiles represent static data, while the last one considers dynamic information.

Since the context is shared in the User Environment, each Home-Box will only maintain one instance of User Profile, which includes all users, devices, services and dynamic information within the home network.

The « User » sub-profile stores the basic identity information of home users (name, age, gender, address, phone number, profession, etc.) and the user preferences. It is composed of one or more "User" elements, whose structure is detailed in **Table 4**.

Table 4: UserType Schema for User Sub-Profile

Element Name	Cardinality	Element Type	Comments
UserID *	1	ID	Unique Identifier for the user (*: Attribute)
General	1	mpeg7:PersonTyp e	Including User's name, affiliation, citizenship, Address, Electronic Address
Activity	0…∞	mpeg7:SemanticB asicType	Usage history
Preference	0…∞	mpeg7:UserPrefer enceType	User preferences
Favorite	0…∞	mpeg7:SemanticB asicType	List of favourite items: books, movies, web pages, etc.
GUI Preference	0…1	GUI PreferenceType	List of favourite graphical appearance details: font, colour, background, etc.

The « Device » sub-profile (**Table 5**) is based on MPEG-21 DIA UED [12] and UPnP Device Descriptions [13]. UPnP device schema is used to describe some hardware information such as serial number and model description of the terminal devices. MPEG-21 DIA provides tools to describe terminal capability, including encoding and decoding capabilities, device characteristics such as power, storage, data I/O, display and audio output capabilities.

Table 5: DeviceType for Device Sub-Profile

Element Name	Card	Element Type	Comments
DeviceID*	1	ID	Unique ID of the terminal device (*: Attribute)
DeviceDescription	1	upnpd:deviceType	Device Type defined by UPnP, including manufacturer, model and associated services.
NetworkInterface	0…∞	NetworkInterfaceType	This type is extended from dia:NetworkType; it describes the network interface(s) of the device: MAC/IP address, interface class (wired/wireless), max/min capacity, error correction capability.
Terminal	0…1	dia: TerminalType	The elements contain capability information on the terminal device: different capability definitions in dia or mpeg7 can be used.
Terminal Capability Definitions that can be contained by Terminal element (dia:TerminalType)			
SupportedFormat	0…∞	mpeg7:MediaFormatType	The media formats supported by the device, including information such as file format, coding format, file size, bit rate, etc.
DisplayCapability	0…1	dia:DisplayCapabilityType	mode, screen size, rendering format, etc.
AudioOutputCapability	0…1	dia:AudioOutputCapabilityType	Frequency, SNR, power, number of channels, mode, etc.
PowerCharacteristics	0…1	dia:PowerCharacteristicsType	Average power consumption, battery mode, battery remaining capacity and time.
StorageCharacteristics	0…1	dia:StorageCharacteristicsType	I/O transfer rate, size, writable or not.

Here is a concrete example of a Terminal element with different capability definition:

```
<ALICANTE:UserProfilexmlns:dia="urn:mpeg:mpeg21:2003:01-DIA-
NS"
xsi:schemaLocation="http://www.ict-
ALICANTE.eu/schema/2011/userprofile
http://www.ict-ALICANTE.eu/schema/2011/userprofile"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:ALICANTE="http://www.ict-
ALICANTE.eu/schema/2011/userprofile"
xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001">
<ALICANTE:Device>
<ALICANTE:Terminal>
<dia:TerminalCapabilityxsi:type="dia:DisplaysType">
<dia:Display>
<dia:DisplayCapabilityxsi:type="dia:DisplayCapabilityType">
<dia:Mode>
<dia:Resolutionhorizontal="1920"vertical="1080"/>
</dia:Mode>
<dia:Mode>
<dia:Resolutionhorizontal="800"vertical="600"/>
</dia:Mode>
<dia:ScreenSizehorizontal="1920"vertical="1080"/>
</dia:DisplayCapability>
</dia:Display>
</dia:TerminalCapability>
<dia:TerminalCapabilityxsi:type="dia:CodecCapabilitiesType">
<dia:Decodingxsi:type="dia:VideoCapabilitiesType">
<dia:Format
href="urn:mpeg:mpeg7:cs:VisualCodingFormatCS:2001:3.1.2">
<mpeg7:Namexml:lang="en">
                                 MPEG-4 Visual Simple Profile
@ Level 1
</mpeg7:Name>
</dia:Format>
</dia:Decoding>
</dia:TerminalCapability>
</ALICANTE:Terminal>
</ALICANTE:Device>
</ALICANTE:UserProfile>
```

The « Service » sub-profile (**Table 6**) is used to record the information about services which the EUs subscribed to (name, version, related protocols and ports, etc.) and the contents that they manipulate (multimedia contents, databases, files, etc.). It also contains information that can serve in the publishing, discovering, presenting and billing phases.

MPEG-7 schema and MPEG-21 DIA UED are used to describe « Service » sub profile as follows:

- *mpeg7:ControlledTermUseType* and the corresponding Classification Scheme are used to define ServiceType and ServiceLevel. They provide a mechanism for defining vocabularies for specific domains in an interoperable and extensible way. A Classification Scheme defines a set of terms, which are referenced in the controlled terms;

- Controlled terms are represented as elements of type *mpeg7:ControlledTermUseType*;
- Classification Scheme is represented as *mpeg7:ClassificationScheme*. The *mpeg7:ClassificationScheme* specifies a domain, to which it is applicable, and it associates a URI with each term it defines. The URI is then used by the *mpeg7:ControlledTermUseType* to reference that term;
- *mpeg7:CreationInformationType* is used to define ServiceDescription and ContentMetaData;
- *dia:LimitConstraintType* is used to define the child elements of ServiceRequirements.

Table 6: Service Sub-profile Schema

Element Name	Cardinality	Element Type	Comments
ServiceId	1	anyURI	Unique identifier for the service
ServiceInformation	0..∞	ServiceInformationType	Complex type including information about the service (name, type, level discovery location, version, period of validity and description)
ContentMetadata	0..1	mpeg7:CreationInformationType	Complex type for describing the service and its associated content
ServiceRequirements	0..1	ServiceRequirementsType	Complex type including information about minimum network and EU display requirement for the service. Other requirements specific for the service can be added.

Figure 7 presents the *ServiceType* elements and **Figure 8** details the structure of the *ServiceInformationType*.

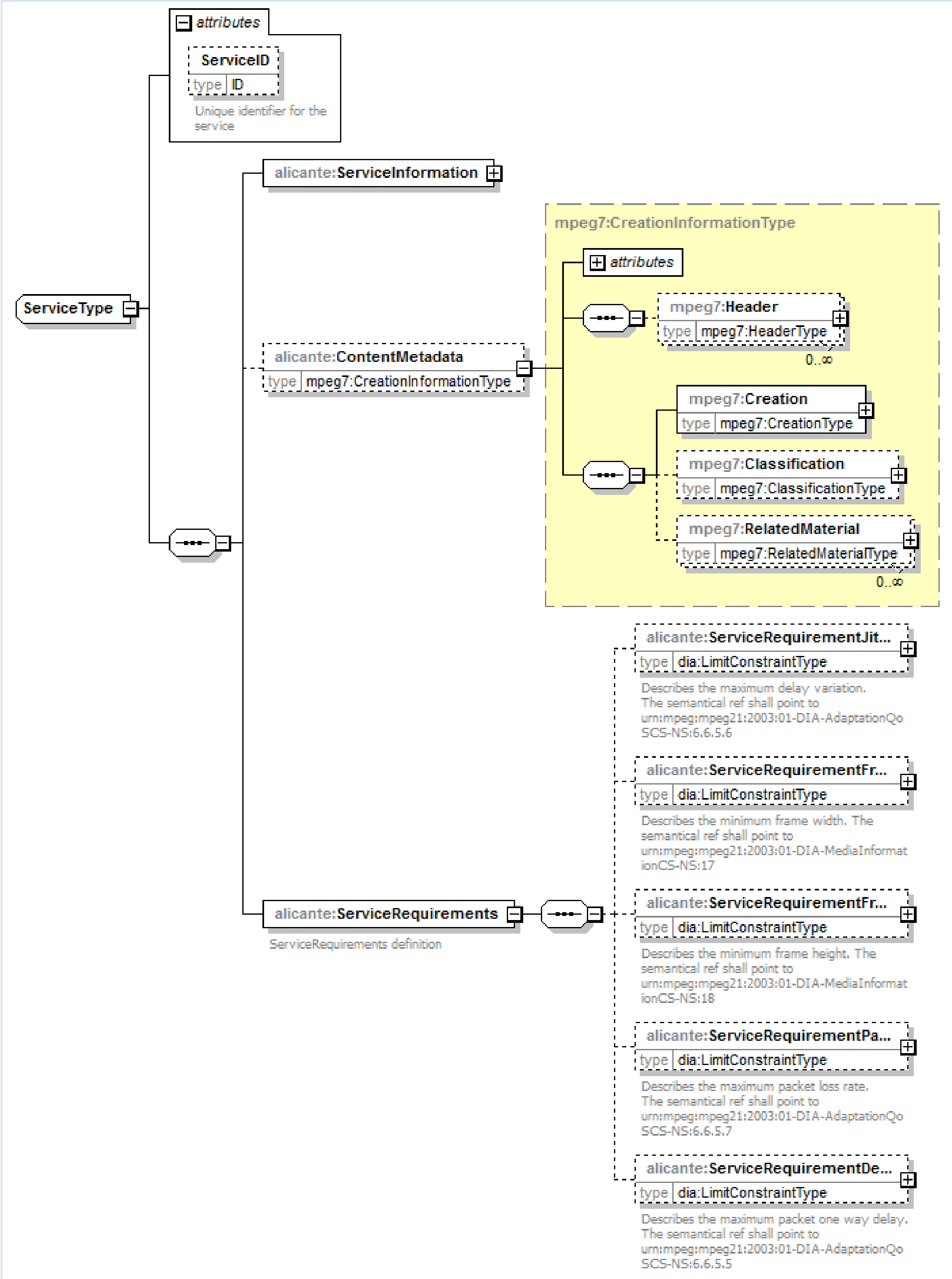


Figure 7: ServiceType Structure

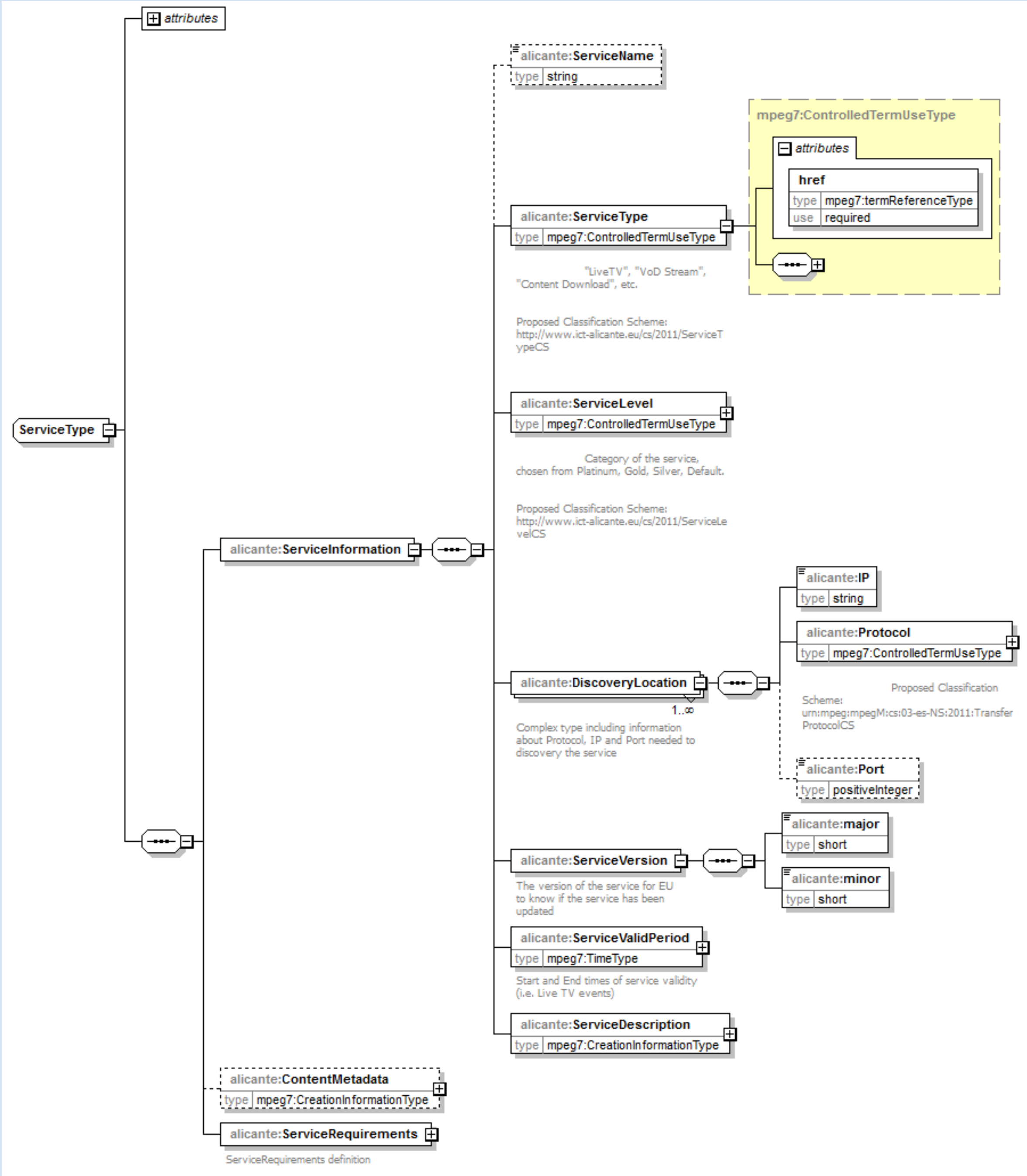


Figure 8: ServiceInformationType structure

The "Context" sub-profile is the dynamic part of the User Profile, which is mainly composed of volatile data that characterize the current User Environment (time, date and indoor/ outdoor location, running applications, etc.), the serving devices (CPU charging, battery level, etc.) and the actual serving networks conditions (available bandwidth, loss and error rate, etc.). The data contained in this part of the profile are collected by monitoring functions embedded at the user terminals.

As presented in **Figure 9**, the “Context” sub-profile is modelled in element Context within a complex type *ContextType* composed of two elements:

- The Session element that references four elements by IDREF attributes: the *UserRef*, the *DeviceRef*, the *ServiceRef* and the *NetworkRef*. These referenced elements are the entities that the session implies (the user that initiates the session, the requested service, the devices and network through which the service is accessed);
- The Generic element *ContextParameter* within a complex type *ContextParameterType* is characterized by Name*,* Valueand Category. A *ContextParameter* is associated to a given session, directly, by referencing it by the IDREF attribute *SessionRef* or, indirectly, by referencing an element already referenced by the session. Indeed, context parameters can be common to more than one session. For example, several sessions can share the same network, so the parameter *NetworkInterfaceUtilisation* is associated with a given network but it is common to all the sessions that share this network.

All the monitored QoS/QoE parameters are modeled as ContextParameter elements. For example, session-related parameters will be modeled as session data rate; QoS parameters as packet loss; Terminal context as CPU load, battery level, etc. Since the “Context” sub-profile is composed of monitored data, each ContextParameter is valid for a period of time and its expiration is specified in the attribute *expiresAt*.

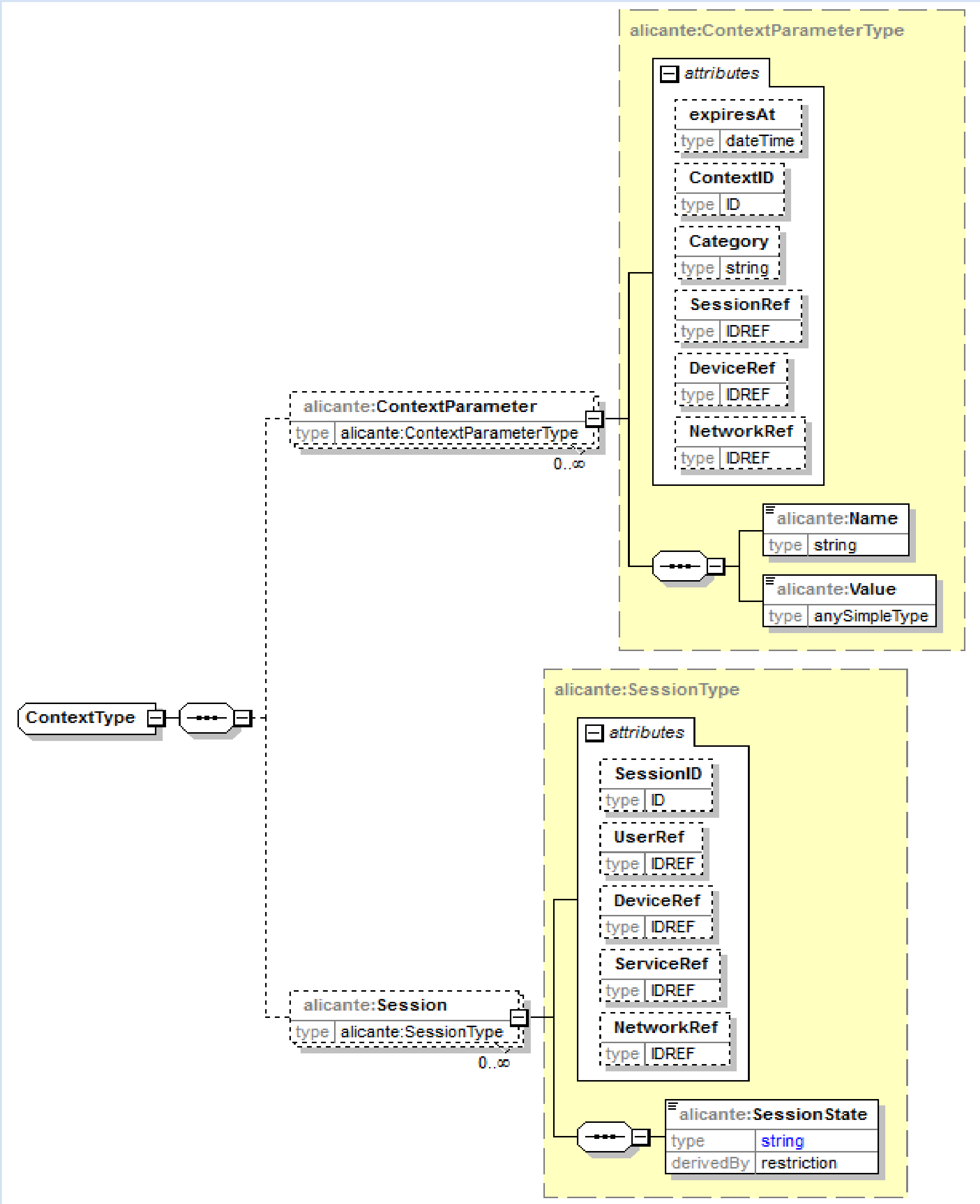


Figure 9: ContextType structure

Based on the context model and in collaboration with the other service management and adaptation components, the Context-Awareness Middleware (CAM) ensures an ubiquitous access to services for the related home users. As illustrated in **Figure 10**, the CAM consists of an intermediate entity between the Context Providers and the Context Consumers. Indeed, raw context data is first acquired from multiple sources located at the different User, Service and Network Environments, then integrated to the context model on which new high level context is deduced thanks to intelligent queries and, finally, provided in the right format to the Service Environment that will take decisions on it and perform some actions to adapt services and contents to the user context.

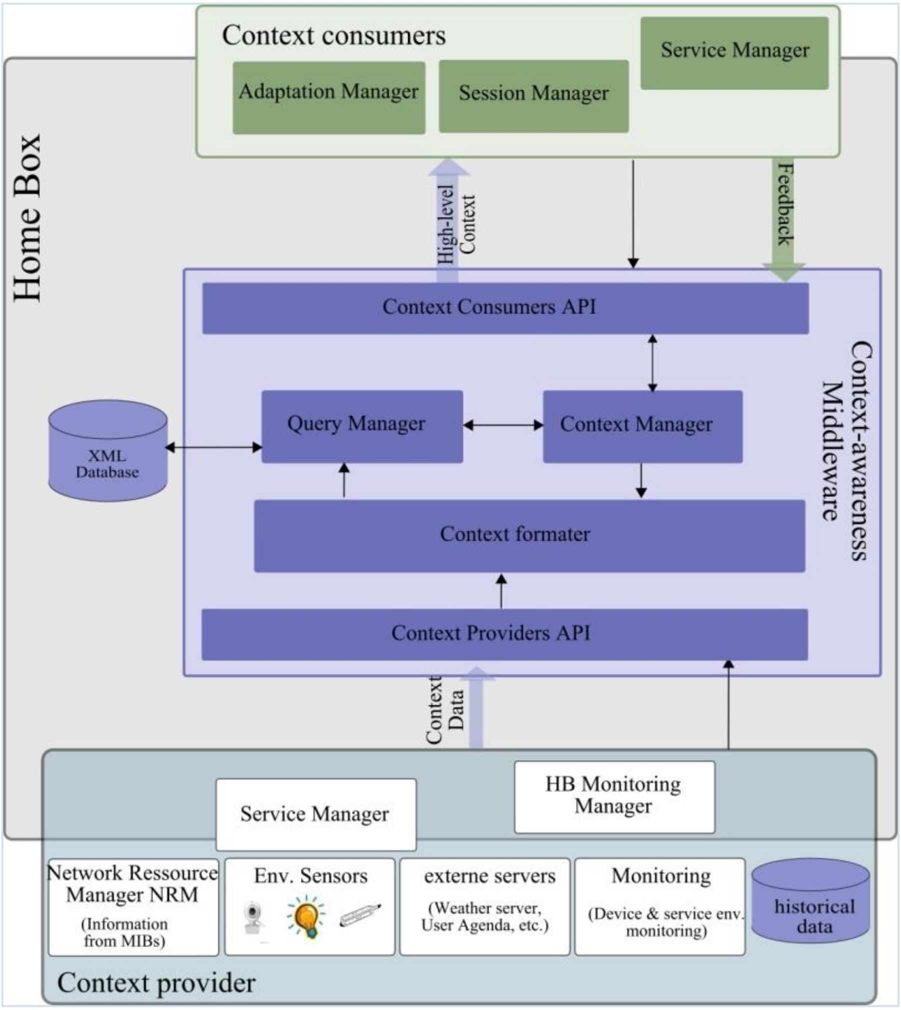


Figure 10: The Context-Awareness Middleware Architecture for Markup

The CAM presents two types of *APIs* for both Context Providers and Context Consumers. A new Context Provider can hence easily subscribe to the CAM to provide it with the context information. Context Consumers can also subscribe to be notified when some context information change.

The *Context Formatter* is in charge of the translation of context information from the format given by the Context Provider to the format defined in the profile schema. In order to enable Context-Awareness, each service should be able to easily access the profile data. This access is managed by the *Query Management* component. It maintains XQuery models that express the context information to which Context Consumers have subscribed. The Query Management is also responsible for the update of the profile stored in the XML database.

The *XML Database* is responsible for the storage of the profile information. Moreover, it offers to other modules the possibility to track important profile parameters using XQuery triggers and to propagate the important events to other components. Both relational database and native XML database can be used. However, since the native XML databases provide better XML support and flexibility [14], we have opted for this choice.

The *Context Manager* is the central component of the CAM. It is in charge of maintaining all the information concerning the Context Providers and Context Consumers. For Context Consumers, it maintains the context information for which they have subscribed, the format (data structure and type) and the mode (periodic/on-demand) in which the latter should be sent and the frequency of context provisioning, if this process is done periodically. In parallel, it maintains, for the Context Providers, the set of information that they provide, the format (if it does not correspond to the profile schema), the frequency and the related access information.

The CAM is implemented in Python. The profile is stored in the native XML database Sedna XML [15] and is accessed through XQuery queries [16]. Sedna XML database is an open source implementation providing a full range of database services for XML native data. Basic operations such as XQuery-based request, update and triggers configuration are supported by the Sedna API.

Concerning triggers configuration, the database trigger is a procedural code that is automatically executed in response to some predefined conditions. It is mostly used to log events, prevent changes and keep the database integrity. Sedna supports XQuery trigger, with the following syntax:

```
CREATE TRIGGER trigger-name
( BEFORE | AFTER ) (INSERT | DELETE | REPLACE)
ON path
( FOR EACH NODE | FOR EACH STATEMENT )
DO {
  Update-statement ($NEW, $OLD, $WHERE);
            . . .
  Update-statement ($NEW, $OLD, $WHERE);
  XQuery-statement ($NEW, $OLD, $WHERE);
}
```

The XQuery trigger can be used to monitor values of QoE/QoS parameters. The following trigger example raises an error when the MOSScore value is lower than 3:

```
CREATE TRIGGER "mos-alarm"
AFTER REPLACE
ON
doc("userprofile.xml")/UserProfile/Context/ContextParameter/
FOR EACH NODE
DO {
if(($NEW/name = "MOSScore") and ($NEW/value < 3))
then error("Low MOS value detected.");
}
```

The communication protocol between the CAM and the other component is SOAP/XML. The soaplib v is used. Python 2.7 Web services interfaces are integrated with the Sednadababase, and soaplib 0.8.1 is used as SOAP server, and suds 0.4, as soap client.

The following SOAP message illustrates an example of message sent by QoE Monitor, located at the user terminal, to the CAM:

```
<?xmlversion="1.0"?>
<soap:Envelope
xmlns:soap="http://www.w3.org/2003/05/soap-envelope">

<soap:Bodyxmlns:m="http://www.ict-
ALICANTE.eu/schema/2011/UPMngr">
    <m:UpdateUserProfile>
           <SessionID>uuid</SessionID>
           <UserProfileParameters>
                  <!-- Parameters to be updated -->
           </UserProfileParameters>
    </m:UpdateUserProfile>
</soap:Body>
</soap:Envelope>
```

The SOAP message is received by the CAM that updates the profile in the database through XQuery Update query. The following XQuery Update query updates the context parameters for a given session:

```
UPDATE
replace
doc("userprofile.xml")/Context/ContextParameter[SessionID=uui
d]
with NEW_CONTEXT_PARAMETERS
```

To conclude, the markup scheme-based context framework provides a light weight XML-based context-aware framework for Media Services. However, the use of more formal models such as ontology and rules can enrich significantly the context expressiveness and the framework flexibility, extensibility and interoperability.

3.2.4 Ontology-based Context Framework

In this contribution [17], it is proposed (1) an ontology-based model characterizing a variety of context information with explicit semantic representation and (2) a large-scale framework that enables context-aware Internet Media Services, in which the context management and adaptation features are distributed among Home-Boxes. For this, we introduce a reasoning-based middleware that is designed to support different tasks involved in the context management feature. The middleware acquires context from different sources, integrates it to the context knowledge database, reasons on it to derive situational context needed to manage services and, finally, supplies the Service Environment with accurate context information needed in the discovery, invocation and adaptation features.

As a formal representation of context is a condition to process it by software agents, a well designed context model is the fundament of any context-aware system. In the following, we first introduce the languages that we have used in context modeling and then present the proposed context model for Internet Media Services.

Developed first in Artificial Intelligence to facilitate knowledge sharing and reuse, ontology may play a major role in Context-Awareness systems. An often cited definition of ontology by Tom Gruber [18] defines it as *a formal explicit specification of a shared conceptualization of a domain of interest*. Ontology provides means that allow a formal description of the semantic of context information in terms of concepts and roles. Context could then be encoded in such a way that software agents should not only process but also understand and draw new conclusions on the user situation through reasoning, as humans could do. For all these reasons, we have studied the possibility to base our context model on ontology.

Prominent ontology languages are the standards RDF/RDF Schema [19], OWL (Ontology Web Language) [20]. Since OWL is much more expressive than RDF or RDFS, it is designed as a standard and has the support of a well known and regarded standard organization (W3C); we decided to rely our context model on it. For more interoperability, the service-related context is also modeled using OWL-S [21] and the rules are modeled in DL-safe SWRL [22].

In this contribution, we have kept the same context information that we have identified as relevant for Media Services in the markup scheme-based context framework. However, by using ontology, beyond representing the structure of context information, the model describes semantic of context information and the relations between them.

In our model, context data is also classified as static and dynamic context. Dynamic context is the one related to the session context and experiences frequent changes during the session duration. Note that classifying context data in the static context does not mean that it does not change but only that it usually does not happen to change during the service usage. The static context is used in the service discovery and composition phases for more personalized services, whereas dynamic context is used in the adaptation phase to keep the service quality at a satisfactory threshold.

As illustrated in **Figure 10**, we have considered six generic interrelated entities that we have identified common to any domain, namely the user, device, network, service, environment and session entities. As in the first contribution, the context is classified as either static context or dynamic context. The first five entities are then classified as static context entities and the last one is the entity to which the dynamic context information will be associated. Since in ontology, not only hierarchical relations can be modeled, the device entity is just related to the network entity by the relation connected to and does not incorporate it as in the XML-based model. Each of the latter entities is then described and extended with more additional concepts and relations that are needed to provide personalized and dynamically adaptive Internet Media Services.

The User entity is described in different profiles as follows:

- The *GeneralProfile* that contains general information about the user such as name, age, etc.;
- The *SubscriptionProfile* that contains information on the different services for which the user have subscribed and the services that he may access;
- The *ContactProfile* that contains the contact information of the user such as his address, phone number, SIP URI, etc.;
- The *AffiliationProfile* that contains information about the different organization to which the user is affiliated;
- The *AuthenticationProfile* that contains information that allows the user to be authenticated;
- The *PreferenceProfile* that contains the user-defined preferences or the deduced preferences from usage. The user preferences could be generic and applied to any service or situation or they could target a specific service or context entity and thus be applied only when the latter is involved.

The *Device* entity is described in terms of (1) *HardwarePlatform* that can be modeled in hierarchical way since the *components* can be *atomic* or *composite* and (2) *Softwareplatform* by presenting the User and System softwares that the device runs. As a specialization of application, the multimedia applications could be associated with other software such as the Codecs.

The description of the *Network* entity comprises information such as the name of the network and the theoretical parameters that characterize it. Dynamic parameters are described further in the Session entity.

The *Environment* entity is the union of different parameters. For each of them, we define its name and its value as well as its properties.

The *Service* entity represents the different services that the user can access. The Internet Media Services provided by the Service Provider as well as the “home” services are represented within this entity. The Service entity is modelled in an OWL-S Service Profile that models its IOPEs parameters. It may subscribe for events that will trigger some actions integrated in the applications implementing the service.

The *Session* entity is related to different session parameters that model the dynamic context, related to the different entities involved in the session. Such parameters could be a user location, a loss or error rate experienced by a multimedia session within a related network and reported by monitoring modules, a dynamically evaluated user experience, etc.

In addition to the OWL property characteristics that enable reasoning, the model also include SWRL rules that add more expressivity to the model and allow some content adaptation and service personalization decisions triggering. The personalization and adaptation processes usually need more context abstraction. These processes commonly base the behavior changes decision on high-level context (called also situational context). The latter is derived from the raw-level using DL- and rule-based reasoning context with the objective of triggering context-aware decisions.

As illustrated in **Figure 11**, our context model is then composed on two ontologies: (1) the upper layer ontology that contains these generic entities and their classification, and (2) the multimedia-specific ontology that extends these entities with concepts, relations and rules that allow a better support of Media Services. It should also be mentioned that the model represents the context information as well as the way in which it will be exploited.

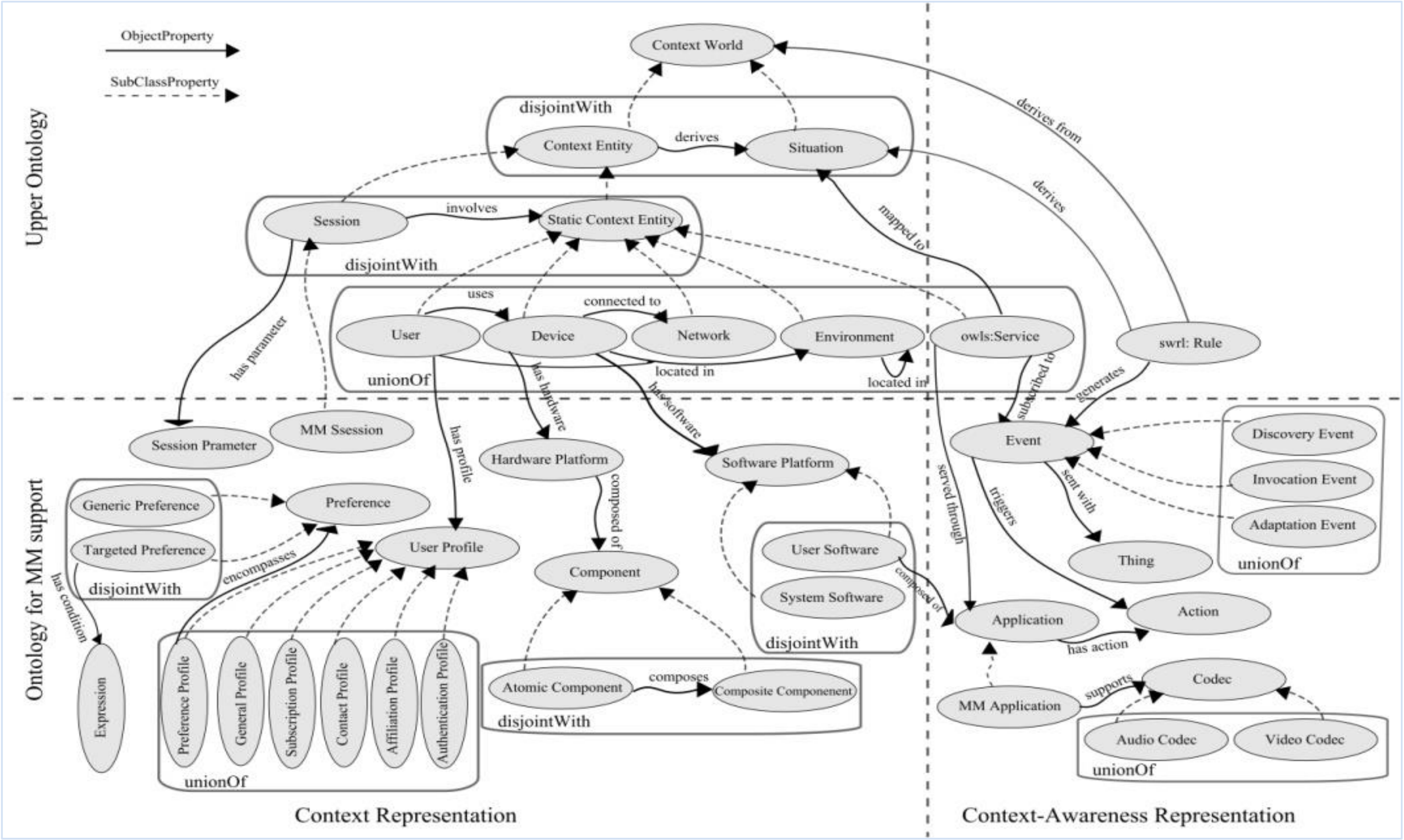


Figure 11: Ontology-Based Context Model

The Service Environment advertises each category of services to Home-Boxes using OWL-S service profile. The service parameters that will be given in the discovery and invocation requests are inferred from the user context and situation using SWRL rules. Discovery and composition tools such as OWLS-MX [23] and OWLS-XPlan [24] that support OWL-S could then be used to retrieve the appropriate service and compose it to fit the user context. In the cases where the service discovery process does not support the Web semantic tools, the service description can then be grounded to WSDL [25].

The following rule selects the category and the size of video parameters according to the user preferences and location. The rule states that when Bob is at his office premises, he prefers to get the match highlights of recently played football matches when he requests for a VoD service:

```
User(Bob) ^ Service(VoD) ^ Environment( ?env) ^
hasPreferenceProfile(Bob, ?pp) ^
encompasses(?pp, ?pref) ^
hasPreferenceName(?pref, Video Category) ^
hasPreferenceValue(?pref, Football) ^
locatedIn(Bob,?env) ^ location(?env, Office) ^
presents(VoD,?service_profile) ^
hasinput(?service_profile, ? param) ^
ServiceParameterName(?param, Video Category) →
sParameter(?param, Football highlights)
```

Rules can also serve to implement some adaptation strategies and/or generate events that will trigger some adaptation processes. The delivered multimedia content can then be dynamically adapted to the context changes. The following rule example triggers a bitrate adaptation of the streamed video content when the network congestion results on a degradation of the service quality. The listener configured on the top of the relation *relatedTo* will send the event to the session manager within the required context parameters in the adaptation process:

```
Service(VoD) ^ Session(?s) ^ Network(?net)
involves(?s, VoD) ^ involves(?s,?net)
hasQoSLossParamThreshold(?serv, ?loss_threshold) ^
hasSessionParam(?s, ?sparam) ^
SessionParameterName(?sparam, registeredLoss) ^
SessionParameterValue(?sparam, ?loss) ^
swrlb:greaterThan(?loss, ?loss_threshold) ^
congested(?net, true) →
relatedTo (?s, AdaptationBrEvent )
```

The rules are maintained by both the Service Provider, to express management and adaptation policies and the End-User, to set some preferences that are related to consumption of the subscribed services.

In addition to the management of the Internet Media Services, the rules can also be used to manage the home services such as the control of the home appliances for more user comfort. The system can then, for example, dim light when the user is consuming video services via their TV or acts on the air-conditioner to prepare the home to the users' arrival according to their schedules.

Based on the context model and in collaboration with the other service management and adaptation components, the Context-Awareness Middleware (CAM) ensures an ubiquitous access to services for the related home users. As illustrated in **Figure 12**, the CAM consists in an intermediate entity/node/agent between the Context Providers and the Context Consumers. Indeed, raw context data is first acquired from multiple sources located at the different User, Service and Network Environments, then integrated to the context model on which new high level context is inferred and finally provided, in the right format, to the Service Environment that will take decisions on it and will perform some actions to adapt services and contents to the user context. The CAM comprises the following functional components:

- *Context Formatter* is in charge of formatting the raw context data, acquired from the different Context Providers, in order to integrate it in the context model. Thanks to this, it permits to abstract the heterogeneity of context data format to the rest of the CAM components. In addition to the validation of context information, the Context Formatter is also in charge of the translation of the user and service policies to the correct DL-Safe SWRL format to be integrated in the rule set of the system;
- *Context Knowledge* encompasses the generic upper layer ontology and the specific one that allows the support of Internet Media Services. It also integrates the management rules that will infer situational context;
- *Context Manager* is responsible for maintaining the context knowledge. It is also in charge of the subscription/notification process that supplies the Context Consumers with needed context information. It thus maintains a set of events that must be generated when the related situations are satisfied and sends them to the subscribed services along with accurate context information needed in the processes triggered by these events. Following the same behaviour as in the previous context framework, the Context Manager is also responsible for maintaining the set of Context Providers that supply the context knowledge through sensing information. For each of them, it maintains its access information, the set of provided context information and its structure, the mode of context provisioning (periodic/on-demand) and its period, if the periodic mode is applied;
- *Inference Engine* reasons on the context knowledge to derive (1) the high level or situational context that will trigger the service and content adaptation actions and (2) the context-aware service profile that will be incorporated to the services request to

enable personalized service discovery and composition. The Inference Engine is based on DL- and SWRL reasoning.

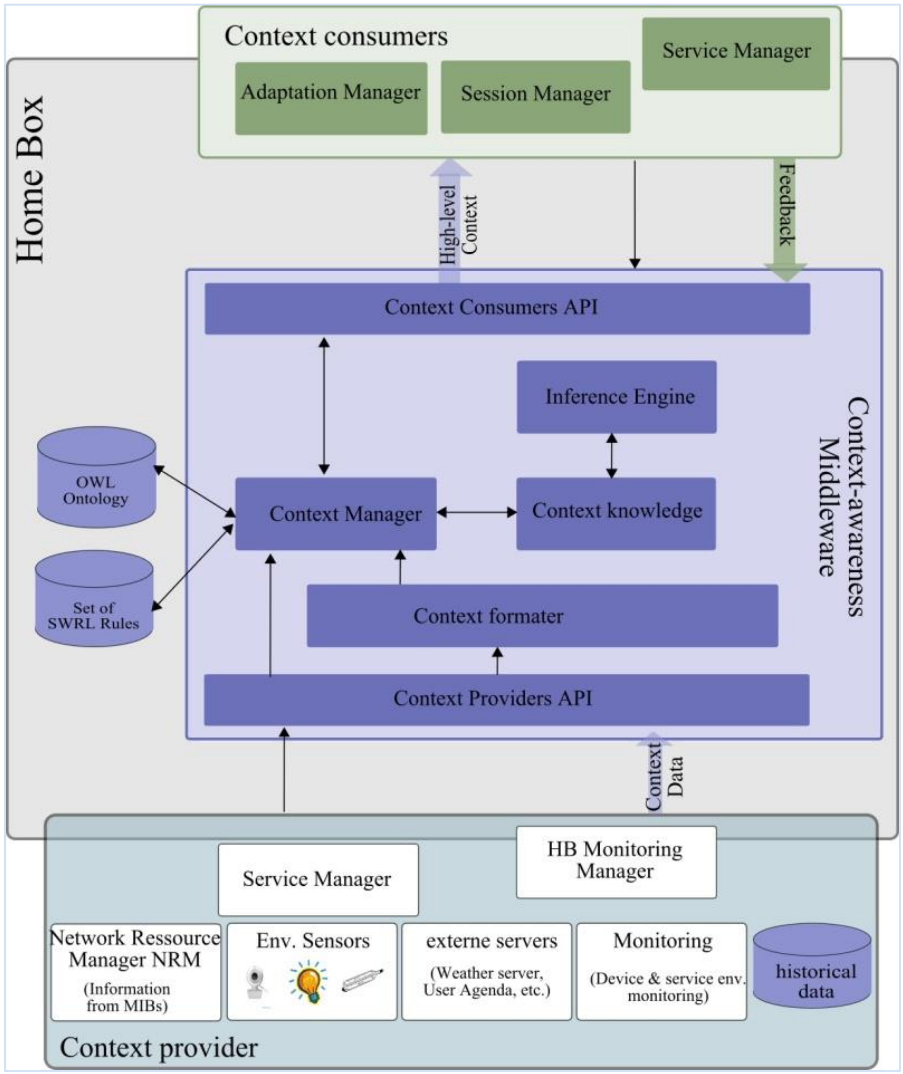


Figure 12: The Context-Awareness Middleware Architecture for Ontologies

A prototype of the proposed CAM was implemented in java 1.6. Protégé-OWL API and the SWRLTab API are used, to manipulate the context ontology and system rules, as well to generate events by configuring listeners and associating to them the context information to be provided with the event. The rules execution is done by the Jess rule engine that is bridged to SWRLTab. The CAM prototype is deployed in the Home-Box component with the following configuration: Intel (R) Core (TM) 2 Duo processor with a frequency of 2.1 GHz and 4GB of memory.

The context ontology is stored as an OWL file at the Home-Box. The ontology comprises the context of 5 users, 7 devices and the description of 3 Media Services (Video on Demand, IPTV and telephony service). The OWL file is composed of 40 OWL classes, 91 properties, 327 individuals and 17 management rules. We have estimated the loading time of the context model to 903 ms and the inference time to approximately 350 ms. The results show that the context reasoning runtime is acceptable for Media Services since the monitoring and adaptation frameworks have commonly longer periods. The loading time is somewhat long but the model is loaded once at the initiation phase.

Concerning the material resources capabilities needed by the CAM, we have derived the following conclusions. The inference process is light in memory since it only consumes about 75 MB of the system memory which is available in all current material configurations. However, it takes around 50% of CPU processing which implies that the current Home-Box has to be enhanced with greater CPU capabilities to be able to handle reasoning-based CAM and supply efficiently very context-sensitive services such as the Media Services.

3.2.5 Conclusions on Context Models

Concluding, the two solutions are able to build a large-scale context-aware framework for Internet Media Services that aim to provide End-Users with such ubiquitous Media Services, always capable to meet their context and situation, as follows:

- The first one is based on the markup-scheme languages that are often used in such a framework for their efficiency. Since XML only supports syntactic modelling, low-level semantic was implemented in the XQuery queries. The markup scheme-based solution is a lightweight solution that is suitable and more performance-effective for multimedia systems. In addition, this solution benefits from the generalization use of XML to reuse some standardized models and related tools. However, it only covers the syntactic level. Although some intelligence and semantic are implemented in the XQuery requests and the context management, it is far to reach the semantic power of ontology;
- The second solution is based on ontology and related languages and tools. The ontology-based solution is richer and allows more expressiveness, flexibility and extensibility. On the other side, it suffers from the complexity of ontology-based reasoning. Such a solution is more resource-hungry. The design decision of distributing the Context-Awareness feature among the Home-Boxes, makes such a solution feasible although its performance still highly dependent on the Home-Box resources. The investments made by ISPs, in the deployment of more powerful Home-Boxes encourages the building of such middleware on top of it.

3.3 The QoE Model

3.3.1 Related Works

In the domain of communication technology, the notion of quality has been associated to the so-called 'Quality of Service' (QoS) for many years. The main motivation towards the recent adoption of the 'Quality of Experience' (QoE) concept has been the basic that QoS is not powerful enough to fully express everything nowadays involved in a communication service.

Hence, in the last years, the term Quality of Experience has been extensively discussed in research literature, normally referring to the user satisfaction during service consumption, i.e., the subjective quality perceived by the user when consuming audio-visual content (usually named Perceived QoS or PQoS) [26]. In [27], the QoE is additionally affected by environmental, psychological, and sociological factors such as user expectations and experience. Recently, [28] proposes to define QoE as "the degree of delight or annoyance of the user of an application or service".

Two basic approaches exist for assessing the QoE:

- **Subjective assessment** as formalized by the ITU-R recommendation BT.500-1 [29], which suggests experimental conditions such as viewing distance and conditions (room lighting, display features, etc.), selection of subjects and test material, assessment and data analysis methods. The tests are performed by employing an evaluator's panel, assessing the quality of a series of short video sequences according to their own personal opinion. The output is the quality of the sequences as seen by an average observer and is usually expressed as a Mean Opinion Score (MOS typically ranging from 1 – bad – to 5 – excellent);
- **Objective assessment** that is being widely adopted in the industry since the preparation and execution of subjective tests is costly and time consuming. The objective evaluation methods involve the use of algorithms and formulas, evaluating the quality in an automatic, quantitative, and repeatable way, based on either signal processing algorithms or network-level quantitative measurements.

Due to the importance of video in today's communication services, we focused the attention to the evaluation of video quality estimation, for which three main methods exist in literature:

- *Full Reference:* to be adopted when the original and processed videos are both available;
- *No Reference:* to be adopted when only the processed video is available;
- *Reduced Reference:* to be adopted when information about the original and processed videos is available, but not the actual video sequences.

Most of the quality metrics proposed in the literature are *Full Reference* metrics, that is, original and distorted videos have to be available to evaluate the quality. These metrics estimate the quality of a video by comparing reference and impaired videos.

Many metrics can be used against subjective analysis; VQEG "Video Quality Experts Group", in conjunction with the ITU-T, published the results as ITU COM9-80-E [30]. The Full Reference Objective Metrics are:

- PSNR -- Peak Signal to Noise Ratio;
- JND -- Just Noticeable Differences;
- SSIM -- Structural SIMilarity;
- VQM -- Video Quality Metric.

The *Reduced-Reference* metrics require only partial information about the reference video. In general, certain features or physical measures are extracted from the reference and transmitted to the receiver as side information to help evaluate the quality of the video. Some examples of *Reduced-Reference* metrics are:

- Objective video quality assessment system based on human perception;
- Local Harmonic Strength (LHS) metric.

The *No-Reference* model does not require any information of the original video sequence; it only makes a distortion analysis of the decoded video sequence to assess its quality and the characteristics of the channel. The main goal of each *No-Reference* approach is to create an estimator based on the proposed features that would predict the *MOS* of human observers, without using the original image or sequence data. And since the model does not require any comparison of signals, the calculations can be performed in near real-time.

As examples of *No-Reference* models, we can cite:

- Blockiness Metrics [31];
- Perceptual video quality assessment based on salient region detection [32];
- MintMOS [33].

From the analysis of the above three group of metrics (*Full-Reference*, *Reduced Reference* and *No-Reference*), it is obvious that *Full Reference* model is only applicable at the server (encoder) side, where the original video sequence is available; as for the *Reduced Reference* model and *No-Reference* model, especially the latter, they are extremely suitable for the wireless and IP video services, where the original reference sequences are absent.

We focused on *No-Reference* methods for the objective assessment and on the *Full Reference* methods for their validation. The validation tests have been conducted in compliance with the

ITU Recommendation BT.500-11 [29], using the Double Stimulus Impairment Scale (DSIS) method. According to the DSIS method, reference content has to be submitted to the audience together to the one encoded by the system under evaluation.

Traditional solutions for objective assessment of quality of real-time services are typically conceived for traditional VoIP and video streaming applications, which are built on UDP as a transport level protocol, and rely on RTP/RTCP for the transfer of real-time data, both favoring timeliness over reliability. Moreover, most of the relevant QoS/QoE models, such as the one described in [34], do not address the problem of evaluating QoE in the case of adaptive bitrate video, i.e., switching among different media representations has in fact an impact on perceived quality.

Due to the recent success of HTTP as a protocol for providing multimedia transmission, the Content Providers are being increasingly interested in evaluating the quality of TCP/HTTP-based services. For such services, the traditional QoS/QoE models are not suitable anymore, due to the different transmission model. For example, QoS parameters such as packet loss rate and packet delay do not apply to TCP-based services. On the contrary, parameters such as buffer underflow/overflow, filling rate, initial delay, etc., have to be taken into account. Withthe exception of very few public works such as [35], the literature on QoE for HTTP based media streaming is still in its infancy. Hence, new challenges face the QoE monitoring in order to address the TCP/HTTP-based services.

Our efforts in the field of QoS/QoE assessment have been concentrated on the development of new QoE models particularly suited for HTTP streaming (DASH, specifically), based on the collection and combination of TCP and media QoS parameters, in order to automatically assess the quality. We also proposed a RTP-based solution for video streaming and for VoIP, as well.

3.3.2 Architecture of the QoE Monitoring Solution

The functional diagram of the QoE monitoring solution is presented in **Figure 13**.

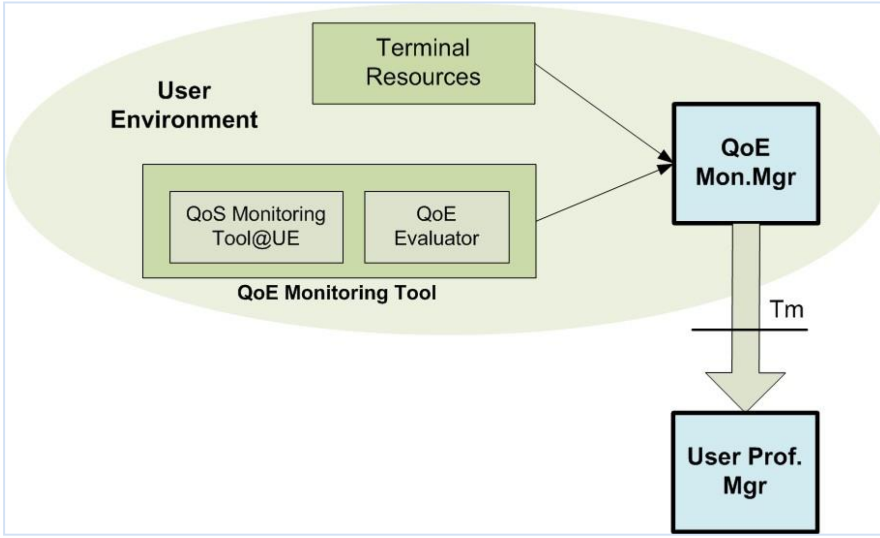


Figure 13: Functional Diagram of the QoE Monitoring Solution

The coordinating module of the whole subsystem solution is the QoE Monitoring Manager (QoEMonMgr), a Python-based software module installed in the Terminal, continuously running in the background in a daemon-like approach, totally transparent to the user.

The QoEMonMgr interfaces with:

- QoS Monitoring Tool @UE, to retrieve network-, session- and media -level metrics for the multimedia sessions presented at the Terminal;
- QoE Evaluator, to retrieve the subjective MOS score for the multimedia sessions presented at the Terminal;
- Terminal OS, to retrieve operational parameters such as CPU load, battery status, interface utilisation etc.;
- User Profile Manager, to store and update the derived parameters, as portion of the Dynamic Part of the User Profile.

User Layer Metrics monitored in our solution are:

- Terminal metrics, which describe the status/resources/capabilities of the Terminal and the media presentation software (i.e., media player); these include:
  - CPU load;
  - Memory utilization;
  - Network interface nominal capacity and utilization;
  - Static system information (such as screen size and resolution);

- Session/QoS metrics, which represent the status of the media stream and of the network; these include:
  - Session state and buffer size;
  - Buffer overflow/underflow events;
  - RTP packet loss;
- Perceived QoS metrics (expressed as MOS), representing the user's perception of the presented service; these include:
  - Audio and video MOS;
  - I/P/B frame loss;
  - Video blur/jerkiness/ringing/noise.

A further refinement has been done in order to make the model more suitable to the MPEG-DASH standard. In fact, traditional solutions for real-time media transmission services are mainly conceived for VoIP and video streaming applications, which are based on the UDP/RTP transport model. Moreover, switching among different media representations has, in fact, an impact on perceived quality.

The general description of the QoE Monitoring Tool is presented in **Figure 14**. The QoE Evaluators has psychometric and PSQA models for RTP streaming. The psychometric QoS/QoE model has been updated in order to better suit the MPEG-DASH scenario.

The QoE Monitoring Tool for DASH is still conceived as the combination of:

- *QoS Monitoring Tool*, with the role of collecting the objective metrics; it is basically a lighter version of a MPEG-compliant media player – e.g., with partial media decoding and no displaying functionalities - including system-level modules for session initiation and control (RTSP/DASH Client, Session Setup), transport-level protocols (RTP/DASH), audio/video decoders;
- *pQoS/QoE Evaluator*, which implements the pQoS model through the combination of the previously collected objective metrics. The QoE is obtained at a later stage through the combination of pQoS and user context information. Such a combination is physically realized in the Home-Box, and particularly in the User Profile. Two models are integrated in the pQoS/QoE Evaluator, but just the psychometric model embeds the features necessary to support MPEG DASH.

The integration at system level of the MPEG stack allows having direct access to the metrics collected by a series of software monitoring probes (Network and Media Probes in **Figure 14**), placed at the media (audio, video codecs), transport (TCP/HTTP, buffering) and DASH (MPD) levels. Such monitoring probes are software APIs functions, which allow collecting the quality metrics at different points of the media playing chain: e.g., at DASH level, a specific API is provided by the libdash module [36] to get the buffer level metric, based on the current level of DASH buffer filling.

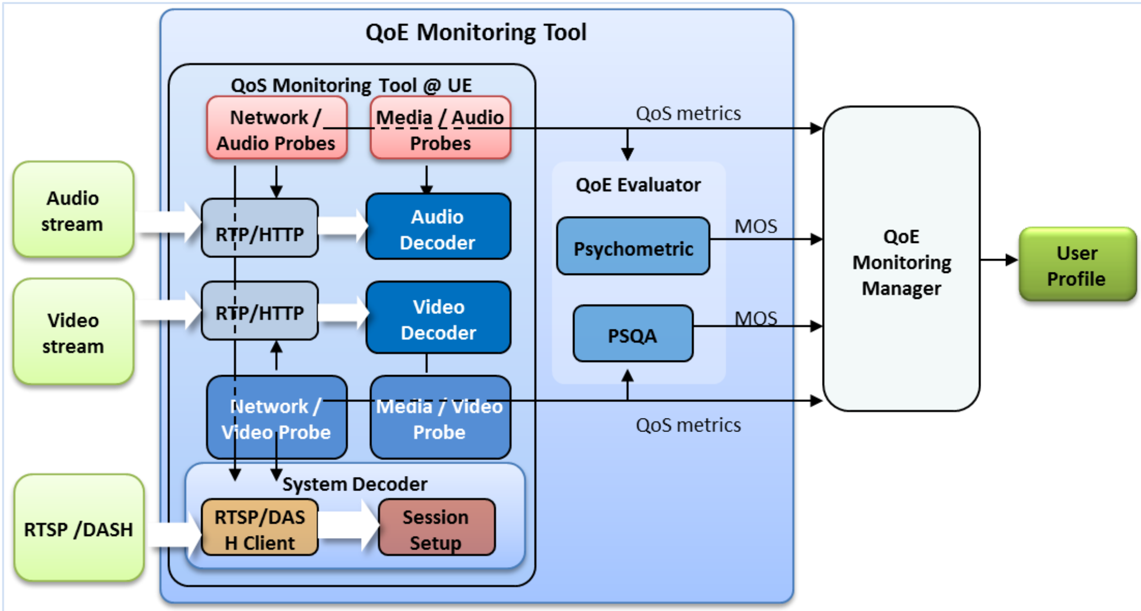


Figure 14: QoE Monitoring Tool Block Diagram

The interface between the QoE Monitoring Tool and the QoE Monitoring Manager, as indicated in **Figure 14**, is used to retrieve QoS / QoE parameters about the multimedia session(s) being presented at the Terminal. Such interface is based on an Inter-Process Communication scheme and, in specific, the Named Pipes mechanism.

The QoE Monitoring tools should be able to estimate QoE score for both RTP and HTTP streams. For HTTP streaming, the DASH standard defines a set of quality metrics for the client able to measure and report back to a reporting server, "should a reporting mechanism be available" [37]. The trigger mechanism is based on the Metrics element in the MPD. The element contains the list of DASH Metrics for which the measurements are desired, the time interval and the granularity for the measurements.

The Metrics can be collected at three different Observation Points (**Figure 15**), as defined in Annex D of the DASH standard [37], which describes the DASH-Metrics client reference model. The Observation Points basically correspond to the three levels (media, transport, MPD) where the monitoring probes, described in the above section, collect the Metrics.

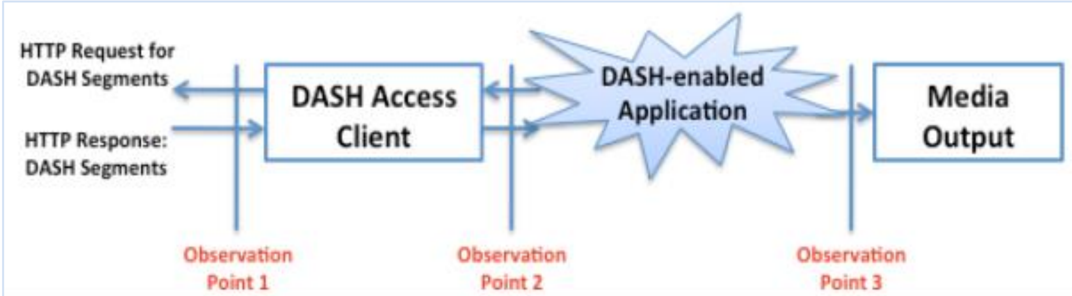


Figure 15: DASH-metrics Client Reference Model

Based on the research done for the design of the above described QoS/QoE model for DASH, we jumped into the discussions within the MPEG DASH group, which led to the Core Experiment named "Signalling of Quality-related information in DASH" [37]. The objective of such Core Experiment is to validate the idea of providing quality information to the client, e.g., concerning sub-segments level for each of the representations within an Adaptation Set. In such scenario, if a client is given information on how the Representation is coded in terms of quality, the clients may benefit. This may allow more intelligent decisions about which Representation to select in order to maintain an overall good quality. At the time while this document is being written, the activities related to such a Core Experiment are still going on and no decision has been taken yet about the possibility of introducing new fields in the DASH standards for transmitting such quality information to the client.

The QoE Monitoring Tool has been conceived as a MPEG-compliant player/sniffer, including multi-standard audio/video decoders and network and system level libraries. Such software solution is composed by a set of C/C++ libraries, supporting:

- Elementary media streams processing (Video: H.264/MPEG-4 AVC and SVC, MPEG-4 part 2 Simple Profile, H.263; Audio: AAC, AMR, G.723, G.729);
- File format reading (MP4, 3GPP, AVI);
- Transport and session protocols (RTSP and RTP/RTCP, MPEG-2 TS, DASH);
- Network sockets (winsock, posix);
- Data flow sniffing tools (libpcap, Winpcap).

The psychometric model has been developed at the system level of the MPEG player in order to have direct access to the metrics collected by a series of software monitoring probes, placed at the media (audio, video codecs), transport (HTTP, buffering), and DASH (MPD) levels.

The data flow sniffing is based on the pcap library [38] (libpcap for Unix-like systems and WinPCap for Windows), which acts at the data-link layer, hence not interfering with the media player being used by the user to play the media content on the same device.

The main application consists of specialized code, depending on the target operating system (Windows, Linux, Windows Mobile, Symbian S60, Symbian UIQ, Android). The audio and video decoding are performed partially, since the entire reconstruction of audio/video samples is not necessary to evaluate the quality parameters; mainly, headers parsing is performed, by extracting the relevant parameters, such as the video quantization parameter, which are all available at the compressed domain.

3.3.3 Evaluation of the Proposed A_PSQA QoE Model

The proposed *Hybrid Quality Assessment* model is based on the so-called "Pseudo-Subjective Quality Assessment" (PSQA) [39]. PQSA is a hybrid assessment method that uses a composition of both subjective and objective evaluation techniques to provide a parametric measure of the QoE. The main idea of the model is to have several distorted samples evaluated subjectively, and then to use the results of this evaluation to teach a RNN (Random Neural Network) the relation between the parameters that cause the distortion and the perceived quality. We slightly modified the PSQA with the necessary parameters for our use, becoming A_PSQA, and we made the evaluations shown below [10].

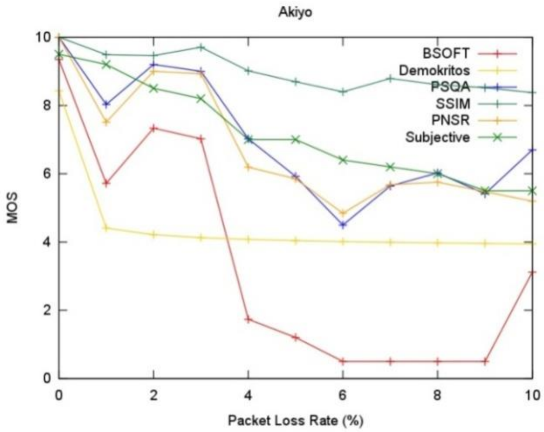
In order to evaluate the proposed model, we used short MPEG-2 video samples (Akiyo, Bus, Football, Foreman and Soccer) with different motion rates (**Table 7**). We coded them to MPEG-2 with the following parameters: Frame-Rate at 30 fps, CIF Format Resolution and a Mean-Bitrate of 300kb. To have a reference MOS score, which allow us to compare the three approaches, a subjective test was conducted with a panel of users using DSIS methods [29].

Table 7: Characteristics of Video Samples

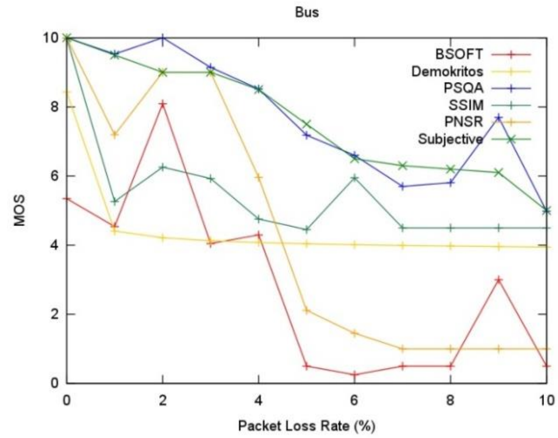
Characteristics	Duration	Motion	Details
Akiyo	12s	Low	Low
Bus	6s	Medium	High
Football	10s	High	Medium
Foreman	12s	Medium	High
Soccer	12s	High	Medium

We used the "Darwin Streaming Server" to stream different samples to the client. To emulate different packet loss percentage, the '*netem*' tool (Random packet loss is specified in the '*tc*' command) was used. The network/video parameters are measured at the client side and then provided to PSQA and other two no-Reference approaches. The received video samples are saved in order to compare them to the original video and to calculate the PSNR and the SSIM index (full-reference methods). Note that, the set of parameters (P) used for training A_PSQA are: I frame loss rate, P frame loss rate, B frame loss rate as well as the Mean Burst Loss Size (MBLS) of each frame type. We considered these parameters as they have the most impact on the Perceived QoE.

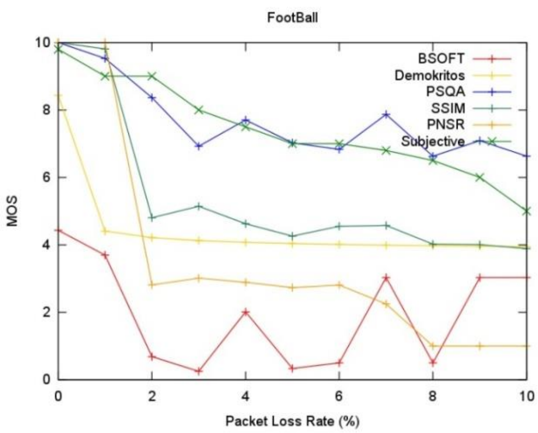
Figure 16 shows the comparison results of different QoE assessment methods. We can notice thatA_PSQA gives MOS (in almost all test cases) very close to the reference value (subjective method), compared to the other approaches. This is true for all sequences with different spatial details and amount of movement (high, medium and slow). A_PSQA is quite robust, comparing to the other approaches, when there is a considerable amount of motion in the video sequence. More information on the A_PSQA and its evaluation can be found in [10] for video streaming service and in [40] for VoIP service. In both cases, A_PSQA shows higher correlation than all the other methods (0,90 for real-time video streaming QoE, and 0,98 for VoIP QoE).



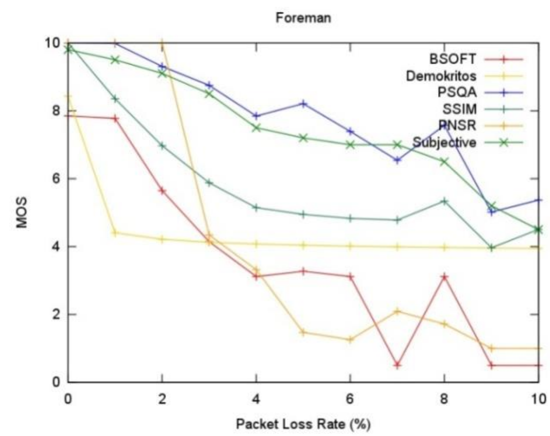
(a) Akiyo



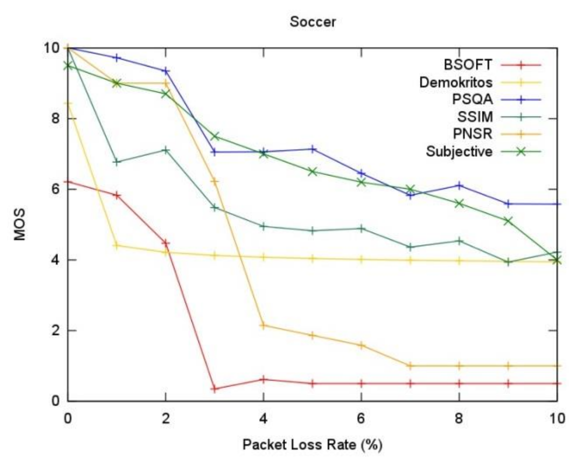
(b) Bus



(c) Football



(d) Foreman



(e) Soccer

Figure 16: MOS Estimation Results with Video Samples

3.3.4 Evaluation of the Psychometric QoE Model for MPEG-DASH

Starting from the psychometric QoE model, we developed a new QoS/QoE model specifically suited for DASH-based services which takes into account subjective quality variations as a trigger of adaptation decisions. The goal is to provide the Adaptation Decision Taking Framework [41] with a significant feedback concerning the current monitored quality, so that it can conveniently react and, at the same time, tune its adaptation strategy. For example, frequent adaptation operations may have a negative impact on the user perception; the new model takes into account the representation switching rate and the steps amplitude in order to measure how the perceptual quality is influenced by adaptation decisions.

At first, we identified a meaningful set of metrics that have an impact on the QoE and, based on these metrics, the QoE model has been established. The new metrics have been categorized as follows:

- *Buffer overflow/underflow metrics*: thresholds in the buffer filling level have to be defined to avoid image freezes or packet losses. TCP is supposed to deliver to the client all the packets, so it does not resend packets in case they are lost at the application level. Re-buffering events are taken into account in the model in terms of frequency and average duration;
- *Frequency and width of quality switches*: this metric is the frequency of the content representation switching, according to a predefined logic, and the related width;
- *Objective content quality level* (media parameters): these are parameters related to the actual media encoding algorithms, providing information on the actual content presented to the user. They are introduced in **Table 8**.

Table 8: Video Parameters affecting the QoE

Parameter	Description	Symbol
Video Bitrate	Indicates the video bit rate. Scale: actual bits per second (bps). Range: 0 to infinity.	BR
Video Framerate	Indicates the video frame rate. Scale: actual frames per second (fps). Range: 0 to 60 fps.	FR
Video Quantization Parameter	Video Quantization Parameter. Scale: average value of MPEG4-AVC QP. Range: 0 to 51.	QP

Together with these objective parameters, the metrics presented in **Table 9**, with impact on the QoE, have been taken into account.

Table 9: DASH Specific Metrics

Parameter	Description	Symbol
Rebuffering Events Frequency (RER)	Indicates the frequency of rebuffering events. Scale: number of events per second (floating point). Range: 0 to infinity.	RER
Rebuffering Events Average Duration (RED)	Indicates the average duration of rebuffering events. Scale: average duration of rebuffering events in the observation period (seconds). Range: 0 to infinity.	RED
Representation Quality Switching Rate (SR)	Indicates the frequency of adaptation switching events. Scale: number of events per second (floating point). Range: 0 to infinity.	RQSR

The algorithm embodying the automated QoE calculation is based on the *psychometric model*; it aims at mimicking and quantifying the psychological reactions of human beings confronted to determined multimedia presentations. The model is based on a non-linear function with the following structure:

$$MOS = \sum_{i=0}^{N-1} a_i * x_i^{k_i} \qquad (1)$$

where:

- $\{x_0 \ldots x_{N-1}\}$ are the measured values of the selected metrics described above;
- $\{a_0 \ldots a_{N-1}\}$ and $\{k_0 \ldots k_{N-1}\}$ are coefficients used to fine-tune the final MOS calculation in order to closely follow the results of subjective testing.

For each x_i, the actual meaning and value range of a_i and k_i will change according to the specific metrics nature.

The $\{a_0 \ldots a_{N-1}\}$ and $\{k_0 \ldots k_{N-1}\}$ values have been determined by submitting the psychometric model to a self-learning process of adjustment of the coefficients and exponents in order to progressively reduce the gap with the scores previously provided through subjective tests, which have been performed according to the ITU Recommendation BT.500-11 [29]. Such an approach has been iterated in order to get the validation of the model based on standard subjective validation procedures. The test procedure has been defined by determining a set of quality affecting factors/parameters, which may have an impact on the perceived quality such as quantization parameter, framerate, re-buffering events, adaptation switching rate, together with the related set of values. A set of parameters is referred to as a configuration; several configurations are defined. Once the configurations have been defined, a set of 'impaired samples' has been built. This means that samples resulting from the transmission of the original media over the network under the different chosen configurations are taken.

Subjective Evaluation

A pool of subjective evaluators has been confronted with the test sequences, in compliance with the ITU Rec. BT.500-13 [29]. The Double Stimulus Impairment Scale (DSIS) method has been selected; the participants had both technical and non-technical background and a medium-high education level.

According to the DSIS method, reference content has to be submitted to the audience together to the one encoded by the system under evaluation. The assessor has been presented with the reference content and then with the impaired one, later during the evaluation session (which lasted 25 minutes) impaired and reference content has been randomly presented to assessors who were required to rate the presented content at regular intervals (1, 5, 10, 15, 20, 25 minutes). At the end of each session, the mean score for each test sequence and impairment combination are calculated among all the subjects.

Objective Evaluation

The objective evaluation consisted in submitting the psychometric model to a self-learning process of adjustment of the coefficients and exponents in order to progressively reduce the gap with the scores provided by the subjective tests.

Figure 17 shows the correlation between MOS values calculated with the psychometric model (PM) and the average values from the panel of subjects (Subj). In an ideal case, the values should be on the diagonal line. The largest divergence between the estimated and the subjective MOS is for the sequence running at 500 kbps, where the estimated MOS is providing lower scores than the subjective MOS. This is due to the fact that subjects are using low quality images for Web-based video streaming as long as the video presentation runs smoothly (no freezes). For the other three sequences, the estimated MOS never diverges for more than 0.4 points and, in some cases they coincide with the average subjective MOS (points on the diagonal line). As an overall, the correlation is very good with a value of 8.55.

These results confirm that the automated calculation of the MOS for a multimedia presentation can effectively rely on the contribution of two classes of metrics:

- Metrics related to the actual media encoding;
- Metrics related to the overall content consumption experience and the events that can impact the End-User.

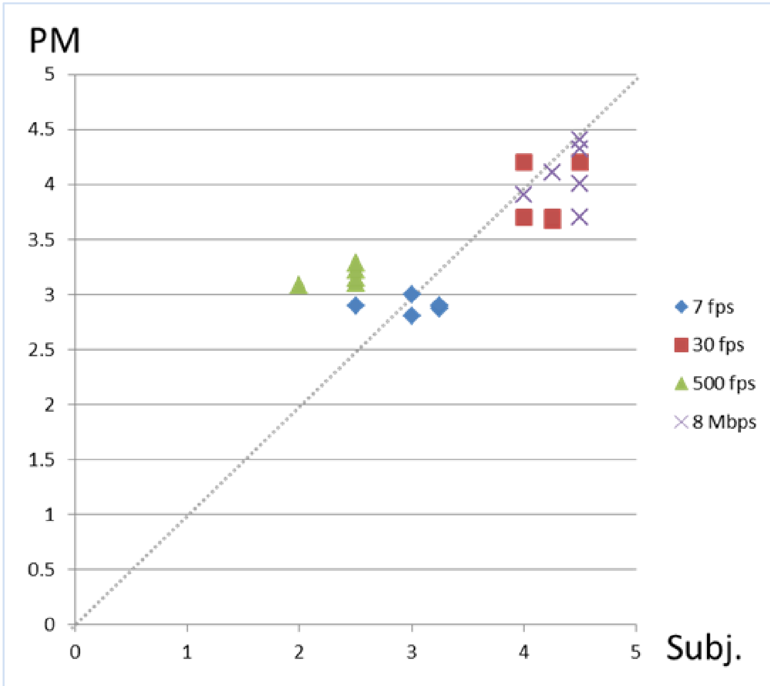


Figure 17: Subjective and Automated MOS Comparison for Video Sequences at 7 and 30 frames/sec. and for 500 kbps and 8 Mbps

These two classes of metrics present fundamental differences, both in terms of measurement and calculation of the impact on the QoE. In fact, while media related parameters affect the presentation at intervals that can be measured in seconds, events such as re-buffering or quality switches will impact the QoE over wider intervals.

In particular, the impact of the representation quality switching rate (RQSR) on the user experience depends on both the actual content quality displayed to the user and to the entity of the quality switch (a wider difference will be more noticeable for the user). Hence, the evaluation of the representation quality switching rate, which is the most DASH-specific metric in our model, required a "recursive" approach where the MOS is calculated based on previous MOS variations, in order to take into account the entity of the quality switch, in addition to the rate. A previous MOS calculation, made at a finer time granularity and, hence, mostly influenced by media parameters, becomes then a contribution to a new MOS calculation, made at a larger time granularity.

The contribution of the RQSR on the automated MOS calculation has been calculated with the following iterative process:

- An observation period is defined (e.g., 2 minutes);
- The quality switches occurrences are recorded together with their width in terms of MOS variation;
- At the end of the observation period, the contribution of each switch event is added to the new RQSR component of the Psychometric Model.

This process is schematized in **Figure 18**.

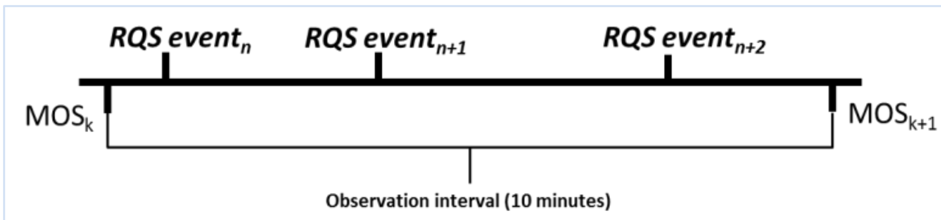


Figure 18: Calculation of Rate Quality Switch (RQS) events on the MOS

Besides the above, it has been demonstrated [30] that the subjective sensibility of the QoE to the same objective metric variation is more pronounced the higher this experienced quality is. If the QoE is very high, a small disturbance will strongly decrease the QoE. On the other hand, if the QoE is already low, a further disturbance is not perceived significantly. We took into account such an aspect by making the QoE variation depending on the current level of QoE, according to a differential relationship, as follows:

$$\frac{\partial QoE}{\partial x} \sim QoE \qquad (2)$$

where, x is a generic objective metric, such as the quantization parameter (QP). The solution to the above differential equation is an exponential function, as the one described in Formula (1). The resulting MOS might be used to enhance the effectiveness of content adaptation policies on a larger time scale, e.g., by dynamically modifying the adaptation reactivity, according to both content and network characteristics and user mobility (which might lead to sudden and significant quality variations). The weighting factors and exponents can be adapted, via off-line configuration, to the service and the terminal targeted by the application, in order to take into account the User Context, as described by the User Profile. For example, re-buffering-related factors can be modified according to the decoder buffer model of the specific player used to play the content in each specific terminal device.

3.3.5 Conclusions on the proposed QoE Model

Within our study, we deployed two models of QoE: one based on PSQA (A_PSQA) for real-time video streaming and VoIP, other one based on the Psychometric model for HTTP (DASH) based streaming. The evaluation of the first one, A_PSQA, shows that it presents the highest correlation with subjective quality rating (MOS) compared to other known quality assessment methods, as well for VOD as for VoIP service. Indeed, the experimental results showed that A_PSQA's correlation scores are very close to 1 (0,90 for real-time video streaming QoE, and 0,98 for VoIP QoE) i.e., there is a very close resemblance between subjective scores with scores given by A_PSQA. This good correlation reflects the robustness and efficiency of A_PSQA. The second one, the psychometric QoE model for DASH streaming, takes into account the property of HTTP streaming (e.g., no packet loss) and estimates the QoE value based on buffer level and receiving bitrate.The results (shown on **Figure 16**) show that there is high correlation (0,85) between the MOS calculated with the psychometric model and the average values given by real subjects.

3.4 Conclusions on the Context-Aware Framework

Context-Awareness is an important trend for ubiquitous computing towards the Future Internet. We propose a **"Context-Aware Framework for Media Services Provisioning"**, trying to answer to different aspects of Context-Awareness, as follows:

- *What pertinent context information to consider?*

The context information depends on the domain of interest. In the case of Media Services, we decided to classify the relevant information in the following categories: 1) User: general description, preferences, activities; 2) Device: screen size, network interfaces, processing capability; 3) Service: metadata, requirements; 4) Dynamic context as session number, available bandwidth, overall QoE. These parameters can be used as criteria for service filtering or as trigger for service adaptation to achieve Context-Awareness. Moreover, these parameters have different importance depending on the service/application. They can be combined in a weighted way to create new parameter. One good example is the Quality of Experience (QoE) which takes into account many aspects such as video, network and user. To acquire pertinent context information, we implemented QoS monitoring tools in charge of measuring QoS parameters, and QoE evaluator to estimate the perceived quality from many objective parameters, using models that correlate adequately.

- *What model to use to formally describe context information?*

The context model is fundamental to the context-aware system. Among all existing models, two types are suitable for User Profile description: markup scheme models (XML) are easy to implement, many tools and standards already exist to facilitate its deployment with existing applications; on the other hand, ontology-based models are also good candidates that offer rich and complete description. We have proposed, implemented and evaluated both User Profile instances in the real context of media adaptation. Both have their advantages and drawbacks, the ontology being more promising in the context of Media even though not yet deployable in the context of large-scale systems.

- *How and where the context information should be managed?*

Depending on context richness and system size, the context information of different End-Users can be managed at different places, in the Service Environment or in the User Environment. We propose a context management system that takes advantage of the Home-Box, which is defined as a "meeting point" of User, Service and Network Environments. In the so-called Home-Box centric architecture, the Home-Box acts as a "context broker" that collects context information from the Home Network, stores the data on the Home-Box and offers an open access to any "Context Consumers" such as Applications or Services Providers, in order to enable the Context-Awareness feature.

Bibliography

[1] M. Weiser, "The computer for the twenty-first century," Scientific American, pp. 92-99, 1991

[2] M. Weiser, "The computer for the twenty-first century," Scientific American, pp. 92-99, 1991

[3] J. Bauer, "Identification and Modeling of Contexts for Different Information Scenarios in Air Traffic', Diplomarbeit, 2003

[4] K. Henricksen, J. Induiska, A.Rakotonirainy, "Generating Context Management Infrastructure from High-Level Context Models", In Industrial Track Proceedings of the 4th International Conference on Mobile Data Management (MDM2003), Melbourne, Australia, 2003

[5] T. A. Halpin, "Information Modeling and Relational Databases: From Conceptual Analysis to Logical Design", Morgan Kaufman Publishers, San Francisco, 2001

[6] T. R. Gruber, "The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases", In J. A. Allen, R. Fikes, & E. Sandewall (Eds.), Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference, Cambridge, MA, pages 601-602, Morgan Kaufmann, 1991

[7] T. R. Gruber, "A Translation Approach to Portable Ontology Specifications", Knowledge Acquisition, 5(2):199-220, 1993

[8] OWL Web Ontology Language Overview, W3C Recommendation 10 February 2004. ttp://www.w3.org/TR/owl-features/

[9] T. Strang, C. Linnhoff-Popien, "A Context Modeling Survey", Proc. of Workshop on Advanced Context Modelling, Reasoning and Management, (UbiComp 2004) - The Sixth International Conference on Ubiquitous Computing. 2004

[10] W. Cherif, A. Ksentini, D. Négru, M. Sidibé, "A_PSQA: Efficient Real-Time Video Streaming QOE Tool in a Future Media Internet Context", proc. IEEE International Conference on Multimedia and Expo (ICME 2011), Barcelona, Spain, 2011

[11] A.K. Dey and Gregory D. Abowd, "Towards a Better Understanding of context and context-awareness", Proceedings of the workshop on the what, who, where, when and how of context-awareness, ACM Press, New York, 2000

[12] ISO/IEC 21000-7 (03-2004), Information Technology - Multimedia Framework (MPEG-21) - Part 7: Digital Item Adaptation

[13] W3C, "Composite Capabilities / Preferences Profile (CC/PP)". http://www.w3.org/Mobile/CCPP

[14] G. Pavlovic-Lazetic, "Native XML databases vs. relational databases in dealing with XML documents", Kragujevac Journal of Mathematics, Citeseer, 2007

[15] Native XML Database System: http://modis.ispras.ru/sedna/

[16] "XQuery 1.0: An XML Query Language (Second Edition)", W3C Recommendation 14 December 2010.http://www.w3.org/TR/xquery/

[17] S. Ait Chellouche, D. Négru, J. Mongay Batalla, "Context-aware multimedia services provisioning in Future Internet using ontology and rules", under review in IEEE PerCom 2014, September 2013.

[18] T. R. Gruber, "The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases", In J. A. Allen, R. Fikes, & E. Sandewall (Eds.), Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference, Cambridge, MA, pages 601-602, Morgan Kaufmann, 1991

[19] "RDF Vocabulary Description Language 1.0: RDF Schema", W3C Recommendation, 10 February 2004. http://www.w3.org/TR/rdf-schema/

[20] OWL Web Ontology Language Overview, W3C Recommendation 10 February 2004. ttp://www.w3.org/TR/owl-features/

[21] OWL-S. Semantic markup for web services; w3c member submission, 22 November 2004. http://www.w3.org/Submission/OWL-S/

[22] I. Horrocks, P.F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, M. Dean,: SWRL: ASemantic Web Rule Language Combining OWL and RuleML, 2004

[23] M. Klusch, B. Fries, K. Sycara, "Automated Semantic Web Service Discovery with OWLS-MX", Proceedings of the 5th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), Hakodate, Japan, 2006

[24] M. Klusch, K-U. Renner, "Dynamic Re-Planning of Composite OWLS Services", Proceedings of 1st IEEE Workshop on Semantic Web Service Composition, Hongkong, China, IEEE CS Press, 2006

[25] Web Services Description Language (WSDL) 1.1, W3C Note 15 March 2001. http://www.w3.org/TR/wsdl

[26] K. Piamrat, C. Viho, J.M. Bonnin, J.M., A. Ksentini, "Quality of experience measurements for video streaming over wireless networks", *Proc. IEEE ITNG '09*, pp. 1184–1189, April 2009.

[27] R. Stankiewicz, P. Cholda, and A. Jajszczyk, "QoX – What is really?", Communications Magazine, IEEE, volume 49, issue 4, pages 148-158, April 2011.

[28] P. Le Callet, S. Möller and A. Perkis, (eds.), "Qualinet White Paper on Definitions of Quality of Experience", European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Version 1.1, June 2012.

[29] Rec. ITU-R BT.500-11, "Methodology for the subjective assessment of the quality of television pictures".

[30] *ITU*-T COM *9-80-E*, "Final report from the video quality experts group (VQEG) on the validation of objective models of video quality assessment", 2003.

[31] S. Winkler, A. Sharma, D. McNally, "Perceptual Video Quality and Blockiness Metrics for Multimedia Streaming Applications", Proceedings of the International Symposium on Wireless Personal Multimedia Communications, 2001.

[32] C. Oprea, I. Pirnog, C. Paleologu, M. Udrea, "Perceptual video quality assessment based on salient region detection", AICT '09, Fifth Advanced International Conference on Telecommunications. 2009.

[33] M. Venkataraman, M. Chatterjee, "MintMOS: Lightweight, Real-Time, no-reference Video-QoE Inference", IEEE Trans. on Multimedia, March 2011.

[34] A. Mishra, "Quality of Experience Based Policy Control Framework for RTP Based Video and Non-Voice Media Sessions", International Journal of Future Generation Communication and Networking, vol. 4, no.2, pp. 81-90, June 2011.

[35] R. K. P. Mok, X. Luo, E. W. W. Chan, R. K. C. Chang, "QDASH: A QoE-aware DASH system," *Proc. ACM Multimedia Systems (MMSys)*, ACM, New York, NY, USA, pp. 11-22, Feb. 2012.

[36] libdash bitmovin open git repository. https://github.com/bitmovin/libdash

[37] A. Giladi, T. Stockhammer (eds), ISO/IEC JTC1/SC29/WG11 MPEG2013/N13286MPEG, Systems subgroup/DASH, "Description of Core Experiments", MPEG meeting output document, January 2013, Geneva, Switzerland.

[38] Official page for tcpdump and libpcap, http://www.tcpdump.org/

[39] G. Rubino, "The PSQA project", INRIA Rennes, www.irisa.fr/armor/lesmembres/Rubino/myPages/psqa.html.

[40] W. Cherif, A. Ksentini, M. Sidibé, D. Négru,"A_PSQA: PESQ-like non-intrusive tool for QoE prediction in VoIP Services", in proc. of the International Conference on

Communications (ICC 2012), Ottawa, Canada, June 2012

[41] M. Grafl, M. Waltl, C. Timmerer, D. Négru, (eds.) et al., ICT **ALICANTE**, Deliverable D2.2, "Service/Content Adaptation Definition and Specification," September 2011.

Chapter 4

The Home-Box-assisted Content Delivery Network System

Global Internet video traffic is growing fast and requires more advanced and scalable distribution schemes. We tackle this need by proposing a novel **content distribution system** relying on a **virtual Home-Box overlay,** exploitable by Service Providers, in replacement or complementarising their existing infrastructures (e.g., CDN), as well as by End-Users. It is a scalable, efficient and flexible solution that **combines current CDN and P2P approaches**, in order to provide **context-aware** features (using the framework presented in Chapter 3), enhanced User Experience (high QoE scores), popularity-adaptive media content distribution, and support for User-Generated Content (UGC) in addition to Video-On-Demand (VOD) services, all considering network-related information (**Network-Awareness** feature).

4.1 Introduction

The wide adoption of Internet in the everyday life along with the growing popularity of Media Services, that are known to be both resource-consuming and performance sensitive, places stringent requirement on current Internet infrastructure in term of scalability and performance. In addition, as the broadband connectivity becomes more and more pervasive, consumers' expectations for personalization and better quality experience in a disparate context increase.

Caching and replication are widely adopted techniques to improve Internet Media Services' performance, scalability and reliability. Their aim is to store copies of contents at strategically chosen locations at the edges of the Internet, closer to End-Users in anticipation of future requests.

The Media Service provision chain in the Future Internet (FI) involves a large number of participants/actors in a "Media Ecosystem": Commercial Providers of contents and services, End-Users acting as both content consumers and creators/providers using various terminals of different capabilities, and Network Operators. Such a "Media Ecosystem" shall enable

QoS/QoE-enabled Media Services sharing and consuming, dynamically exploiting content and resources from the involved actors.

The proposed NetME Architecture in chapter 2 considers this and acts as an efficient solution, taking advantage of the new Home-Box layer caching capabilities in order to improve the services scalability, availability, reliability and access time, towards a better End-User experience. This **Home-Box (HB) layer** enables End-Users (EUs) and Service Providers (SPs) to share and distribute both legacy and novel multimedia-based services. It is composed of virtually interconnected "**Home-Boxes (HB)**", capable of advanced ways of service/content provisioning in various modes at the edges of the multi-domain system.

Based on this architecture, our contribution is oriented on User-Generated Content (UGC) and Video-on-Demand (VoD) services distribution through the Home-Box layer. Focusing on those types of services, we propose a novel **Home-Box Assisted Content Delivery Network System**, including two innovative features: (1) a **Network- & Context-Aware Anycast-based Server/Peer Selection** mechanism that allows End-Users to always access the service from the best server/peer that meets the requirements from the End-Users, the Service Providers and the Network Providers, and (2) **an Online Popularity-Based Video Caching Strategy** in Home-Boxes' caches that efficiently spread the video contents over the HB overlay. Beforehand, we present the related works, and then the description and evaluation of both features.

4.2 Related works

We will first outline related work in the domain of **caching** and replication techniques and then, related to **anycast**. More details and analysis on those two domains can be found in [1] [2].

<u>Caching</u>

Caching can be classified in three categories according to the location of caches in the network [3], namely: the browser cache, proxy cache and surrogate cache. Browser cache is located in the client host as part of the browser to exploit the temporal locality of the user's requests. This type of caching has the least/smallest benefit since the cache is usually quite small and there is no sharing between End-Users. Surrogate caches are located at the Web server side and are typically owned and operated by the Content Provider. The aim is to accelerate the server's performance. Concerning proxy caches, they are located in nodes between the server and the End-User, typically closer to End-Users than to the Content Provider servers. These nodes are owned by Network Providers or by companies operating caching nodes connected to the Internet. This type of caching reduces service access time and permits to save bandwidth by bringing contents close to End-Users. The distinction between the three types of caching techniques is useful from the deployment point of view. However, from research perspective, these cache types share many research challenges. Thus, solutions developed for one type can be directly applied to the others as it will be seen further in this chapter.

The benefits of caching and replication are numerous [3][4][5]:

- From the perspective of network infrastructures, these techniques decrease network traffic and thus minimize network congestions and improve performance. In addition, for Content Providers that pay ISPs to transport their traffic, reduced traffic means lower costs;
- From the Content Providers point of view, caching and replication reduce workload of their servers and service availability, reliability and responsiveness;
- From the client point of view, caching reduces significantly the service access latency for both popular contents since they get them from nearby servers and unpopular contents since the contents are faster retrieved due to reduction of network congestion.

However, there are a number of potential problems related to Web caching and replication. For example, cache misses (when the content is not present in cache and has then to be retrieved from the origin server) decreases the service access time due to the cache processing. Users might also consume stale/out-dated contents, if caches are not properly updated.

Maximizing the benefits of caching and replication solutions requires a careful and intelligent design. Some issues such as cache organization and cooperation, cache placement, decisions on the cachable contents, on when and where to place or replace contents and which cache will provide a requested content from a certain client, have then to be solved [6].

Anycast

Anycast is one of the schemes that support the request redirection. It was originally defined at the IP level [7] as a service that allows a node to connect to one, and maybe the "best" (according to the routing protocol metric), member of a group that serves the anycast address. The group is formed of resources that offer interchangeable services. It can then consist of the surrogates, either servers or proxy caches, which can deliver the requested content.

Although IP-level anycasting was described since almost two decades ago, very few practical networks have been built considering it. This is probably because of the technical challenges inherent in efficiently locating the best server. DNS [8] deployment and "6 to 4" [9] router constitute the most widely publicized uses of unicast addressing scheme. Other research works were made on network-layer anycasting. [10][11] focus on the scalability issue, others such as [12][13] focus on the design of routing algorithms based on active routers; papers such as [14] and [15] proposed proxy-based infrastructures to address network-layer anycast issues, like scalability or session-based services support.

However, this network-layer anycasting approach presents some limitations such as (1) the necessity for routers to support anycast and (2) the allocation of IP address space for anycast address, making difficult its integration on the existing infrastructure. Furthermore, this approach does not consider any user context options, neither the stateless nature of IP nor a set of metrics for choosing the most suitable server.

The network-layer anycasting limitations led the researchers to define the anycast paradigm at the application layer. In [16][17][18], the authors define the anycasting paradigm at the application layer as a service that maps anycast domain names into one or more IP addresses using anycast resolvers. The resolver decides which server, among the replicated ones, is the best based on different metrics at network and server side. For this purpose, the resolver maintains the servers' performance information. Application-layer anycasting consists of a good solution for distributed Internet Media Services provisioning, especially when it requires no modification in the existing infrastructure. Another motivation to use application-layer anycasting is its ability to manage QoS and to define service requirements on a per-service basis.

In the field of Media Services distribution, authors in [19] present an algorithm theoretically related to an economical model with a queuing theory based on the available free buffer, the available bandwidth, the average arrival rate of requests and the call blocking probability. In [20], three anycast-based multimedia distribution architectures are proposed, namely the identical, the heterogeneous and the semi heterogeneous architectures, to identify the best media server selection for different application domains. From our best knowledge, except in [20], all the above works have based their selection strategies on the servers' performance and have considered neither the client context nor the network conditions, which strongly affects the quality of Media Services delivery. Additionally, the solution presented in [20] is only designed and evaluated in small-scale environment.

Our proposed application-layer anycast architecture aims to provide clients with an efficient and transparent Media Services provisioning. As described in chapter 2, we assume that the Service Environment (SE) hosts a Service Registry (SR) node which will receive the requests of its related clients. The SR has two main roles: first, it plays the role of an anycast resolver that is in charge of performing the mapping of the anycast address of the client request to the unicast address of the most appropriate server. Second, it will retrieve and maintain the servers' contexts and their contents descriptions. We also assume that the set of SRs can also collaboratively perform the server selection feature, when the latter fail at one of them.

4.3 A Network- & Context-Aware Anycast-based Server/Peer Selection Mechanism

The Server/(Home-Box) Peer Selection mechanism we propose is based on an **anycast** model and considers Network and Context information. This mechanism is beneficial for Service Providers which are willing to consider making use of a Home-Box overlay architecture, as proposed here, in replacement or in conjunction with an existing infrastructure (e.g., CDNs with their request redirections messages).

4.3.1 The Proposed Network- & Context-Aware Anycast-based Server/Peer Selection Mechanism

The first step of our proposal was to consider video content replication at the Service Environment side [21][22]. Video contents were then replicated among surrogate servers distributed over the Internet. The infrastructure, as illustrated in **Figure 19**, is based on three types of nodes: the client's nodes, in the User Environment, requesting the service, the server's nodes providing the service and the SRs nodes serving the anycast address, handling the client's requests and performing the server selection strategy. **Figure 19** also depicts the engaged dialogue between the different agents that run in the system nodes in order to establish the media session.

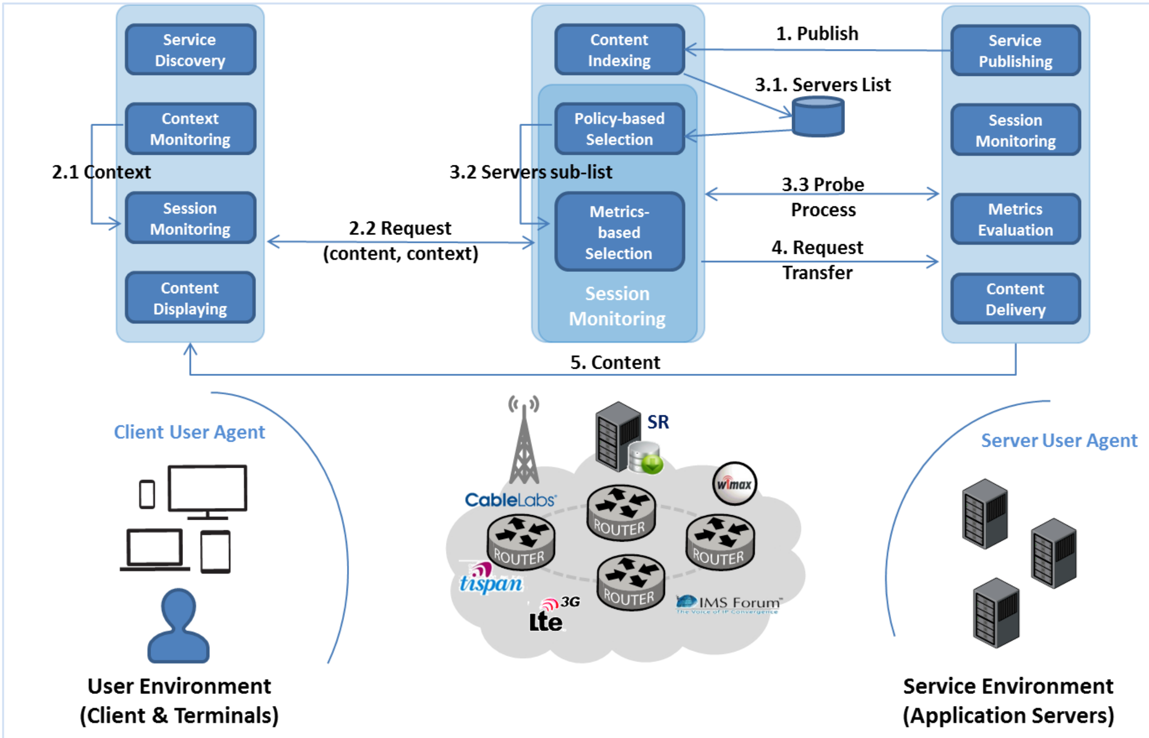


Figure 19: First Anycast-model Functional Architecture

Our proposed approach for video distribution was mainly based on the selection of the best server for each client request among an anycast group of servers. By best server here, we mean the non-overloaded server that best suits both the client environment (connectivity and user & terminal characteristics) and the requirements of the underlying network conditions – from the server to the client – for finally improving the perceived QoE at the client side. To this end, a two-level filtering technique was conceived in order to ensure the accuracy of the selection result: the first level is based on "policy-based filtering" (Context-Awareness), the second level on "metrics-based filtering" (Network-Awareness). The server selection algorithm relies on this 2-step filtering process, as follows:

- **Policy-based filtering**: each SR maintains, at its side, the list of all video contents published by the attached video streaming servers. For each video, it maintains both a set of servers that deliver it and a set of metadata that describe it. When requesting a service, the client specifies, in addition to the requested video reference, its context (available bandwidth, terminal resolution, etc.) and preferences like the video language by including them in the request that initiates the service. The user context is acquired from the context-aware function, as specified in Chapter 3. The registered services, at the SR side, are then filtered by a set of predefined policies in order to only keep the services that deliver contents matching the user context. The policies define the mapping of, on one side, the maintained servers' contexts and related contents information and, on another side, the client context and requirements. The mapping is highly dependent on the model in which the user context and the service and content metadata are represented. At the end of this step, the SR built a list of servers which provide contents that meet the user context, in order to take advantage as much as possible of the context-aware feature;

- **Metrics-based filtering**: the objective of this second phase is to select one server (the best) from the list constructed in the previous step. This selection will be performed based on a **distance** evaluated during the selection process. The metrics to use and their exploitation strategy directly depend on the application and will therefore be selected accordingly. Since we address the video streaming service, which is known to be very sensitive to the packet loss and delay metrics, the main requirement that we have considered when designing our server selection strategy is to avoid congestion and this, at different levels. At the policy-based step, we have considered the congestion at the client level by taking into account the client context and thus the client connectivity. At this step (metrics-based), we consider the congestion at both the server and network levels. Thus, the defined filter for this step is a weighted function that involves two metrics: the server load and the server-to-client delay. The combination of these two metrics permits to avoid congestion (1) at the server side, by avoiding server overloaded situation, and (2) at the network side, by considering the current client-to-server delay. The evaluation of this function is processed as presented in **Figure 20**.

$$F\left(A_c, Rbr_c\right)\{$$

$$d_{sc} = delay\left(A_c\right);$$

$$\underline{if}\left(\left(\sum_{i=1}^{n} Rbr_i + Rbr_c\right) \Big/ br < 1\right) \underline{then}\{$$

$$\underline{return}\; d_{sc};$$

$$\}\underline{else}\{$$

$$\underline{return}\left(\alpha * d_{sc} * \left(\left(\sum_{i=1}^{n} Rbr_i + Rbr_c\right) \Big/ br\right)\right);$$

Figure 20: The Server Selection Function Algorithm

The SR probes all the servers that constitute the retrieved sub-list, from the policy-based filtering, in order to evaluate, for each of them, the server selection function F described in **Figure 20**. Because F combines the server-to-client delay and the server load, the probing request must contain the client address Ac and the required video bitrate Rbrc. On the other side, the server must also maintain its load. Indeed, whenever the server accepts the establishment of a multimedia session or ends one of its current sessions, it must update its load. The current load is calculated as follows:

$$load = \sum_{i=1}^{n} Rbr_i \Big/ br \qquad (1)$$

where:

n is the number of currently active video sessions at the server side ;

Rbr_i is the already required and allocated bitrate for the video session i ;

br is the bitrate of the output link of the server.

The probed servers then evaluate the function F. As illustrated in **Figure 20**, F is based on the server-to-client delay d_{sc} and the server load. While the server is not overloaded, it only takes into account the client-to-server delay. But as soon as the server gets overloaded, the function involves both the two metrics and the priority is inversed. The α parameter should be fixed according to the network topology in order to give the server load metric the top priority.

After receiving the servers' evaluations of the selection function F , the SR selects the best server s such as:

$$F_s(A_c, Rbr_c) = \min_{1<i<m}(F_i(A_C, Rbr_c)) \quad (2)$$

Where, m is the number of the received responses.It should be noticed here that m is not necessarily equal to the number of probed servers. For each client request, the SR sets a timer and when this timer expires, if the SR had not yet received all the responses from the probed servers, it selects the best server based on the received responses.

All the details and results of this first step can be found in [21][22].

One step further, the anycast-based server selection is applied to the HB-based video delivery infrastructure, as shown **Figure 21**. We consider replication at the HB nodes. The HB is enhanced with caching capabilities. We assume that contents are pushed from Content Provider servers to the HB overlay. The way the contents are replicated over the HBs will be detailed later.

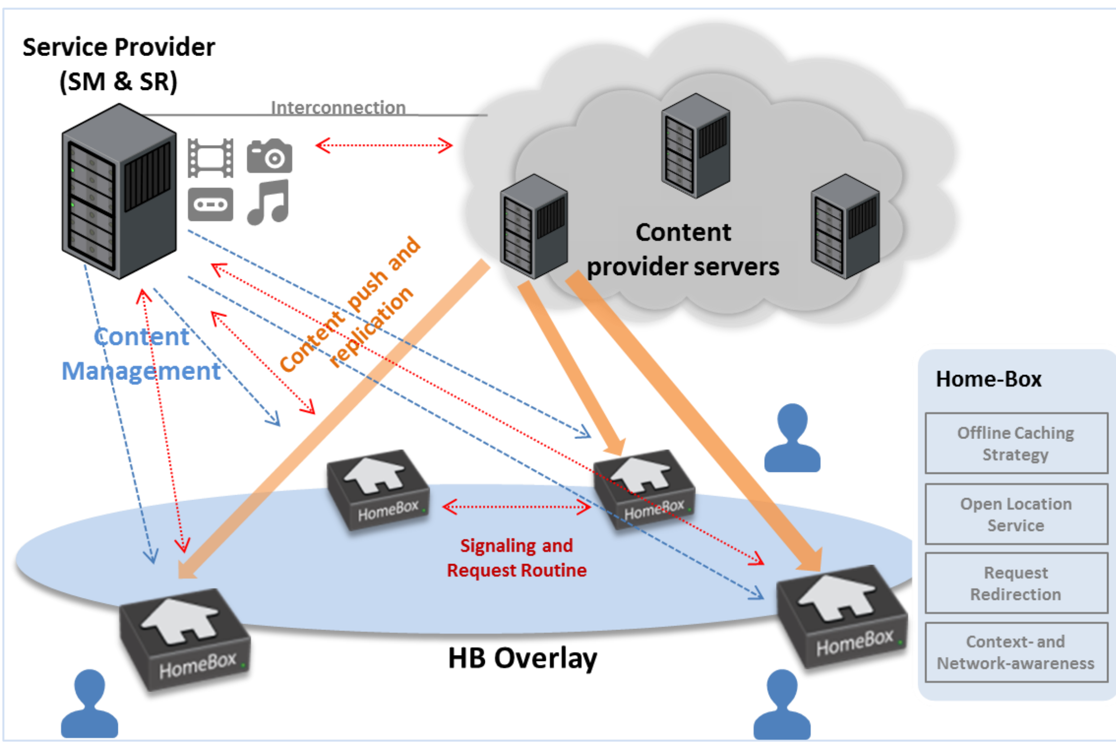


Figure 21: Collaborative Caching Approach with HB Replication

The goal of placing video contents in the Home-Boxes' overlay is to minimize the clients' costs to access them. Indeed, the Home-Box is located at the User Environment which makes it the closest equipment to End-Users that can be managed by SPs. So, considering caching solution on this equipment will significantly decrease latency, save the server and network bandwidth and accordingly, decrease packet loss.

The video contents placement in the HBs caches is performed as detailed below.

Considering N videos, classified in the order of their popularity from 1 to N, only the set of most popular videos Vp are cached in the HBs. Previous works, such as in [24][25], demonstrated that video popularity fits the zipf-like distribution with highly variable skew factor. Relying on the zipf-like distribution, the probability of requesting the video content of rank k is:

$$P_N(k) = \frac{1/_{k^\alpha}}{\sum_{i=1}^{N} 1/_{i^\alpha}} \qquad (3)$$

where, α is the skew of the distribution.
In our model, the video popularity is also reflected in the frequency of the video at the HBs overlay. The frequency of presence of the most popular videos is also evaluated based on the video popularity with the same skew value. Accordingly, the more the video is popular, the more its frequency of presence in caches is bigger. Then, the probability of presence of a video of rank k in the HB caches is:

$$P_{|HB|}(k) = \frac{P_N(k)}{\sum_{j=1}^{M_p} P_N(j)} \qquad (4)$$

The videos are placed in HB caches in the order of their popularity. The most popular are placed first. For simplicity, we assume that the video popularity is user location independent. Thus, for each video, the set of $\left(\mathrm{P}_{|\mathrm{HB}|}(\mathrm{k}) * |\mathrm{HB}|\right)$ HBs that will cache it are selected uniformly from the $\left(\left(\sum_{\mathrm{i}=1}^{\mathrm{k}-1} \mathrm{P}_{|\mathrm{HB}|}(\mathrm{k})\right)| * \mathrm{HB}|\right)$ available HBs, with k being the video popularity rank. Consequently, the most popular videos are then best spread.

Our selection is based on the context-aware anycast server selection strategy explained previously and relies on a two level filtering process. Nevertheless, the selection was adapted to the new architecture that includes the HB virtual layer. Since the content can be accessed from both HBs and CP servers in our model, the term "server" designates both of them.

As detailed in the pseudocode below, the Service Manager selects the closest server, from the resulting list of the policy-based filtering retrieved from the SR, based on accurate metrics retrieved thanks to an on-demand probing process. Each probed server evaluates its load and returns it to the Service Manager. Based on the load information retrieved from the servers themselves and the server-to-requesting HB delay retrieved from the Network Ressource Manager (NRM), the Service Manager evaluates the distance for each server and proceeds to the selection process. The result will be either the closest HB, if the responding HBs are not overloaded, or, otherwise, the closest CP server.

```
/************************************************************
***
    SMngr side pseudo code
*************************************************************
**/

inputs :
L ;  //list of servers provided by the policy-based step
A ;  //address of the local HB
br ;     //required bitrate for the media session

//probe servers to get accurate metrics values
function send_probe_req( L, A, br) {
foreach s ∈ L do {
    send_req(br);
    retrieve_delay(s, A)  /*retrieve from cache if it exists
otherwise from NRM the delay from s
to the client  */
arm timer t;
}
}

// call the select function if the timer expires
function timet_expire(t) {
  //check if the selection hasn't been done
if (!selected) {
selected = true ;
select();
  }
}

/* update the servers context and call the select
function if all the servers have responsed      */
function receive_probe_resp(resp, s) {
if (!selected) {
    L[s].load = resp.load;
nb_resp++;
if (nb_resp == L.size()) then{
selected = true ;
select();
```

```
    }
  } else
    Ignore response;
  }
}

/*select the HB or the Content Server
that will respond the client        */
function select() {
  HBdelay = ∞;
  Sdelay = ∞;
selecteHB = NULL;
selecteS = NULL;
foreach s ∈ L do {
if L[s].type == HB) {
if (L[s].load > 1) then{
L.erase(s);
}else if (L[s].delay < HBdelay) {
        HBdelay = L[s].delay;
selecteHB = s;
      }
    } else {
if (L[s].load > 1) then {
        L[s].delay = α * L[s].delay * L[s].load ;
}
if (L[s].delay < Sdelay) then {
        Sdelay = L[s].delay;
selectedS = s;
      }
}
  }
if (HBdelay < ∞) then {
return selecteHB;
  } else {
return selecteHB;
  }
}

/************************************************************
***
    server side pseudo code
************************************************************
**/
inputs :
  sum_br ; //sum of engaged media sessions bitrates
br ;       //the bitrate of the server uplink

//process the probe request and reurns accurate metric values
```

```
function receive_probe_req(req) {
load = (sum_br + req.br) / br;
  send_probe_resp(load);   /* response SMngr whith the
requested metrics values*/
}
```

An important parameter to measure the video distribution efficiency is the distribution cost. Since getting the video from neighbour HB is almost always less than getting it from a Content Provider server, the aim is then to maximize the HB overlay hit ratio (the ratio of request satisfied by the HB overlay from the total number of requests). Then before deriving the system cost, we will first derive the HB overlay hit ratio $\mathrm{h(R)}$.

Assume C, the number of clusters formed by the HBs and R, the request rate that represents the average arriving requests per T, which is also the time that takes the video to be streamed. R is then, the number of concurrent streams and also the duration for which we will derive the system cost. But first, let us consider the hit ratio $\mathrm{h_R}$ of HBs caches for the R concurrent streams. $\mathrm{h_R}$ is defined as the ratio of video requests satisfied by the HB caching solution.

As explained abovee, the requested video i is accessed from an HB cache only if the video is present at one or more than one HB of the same cluster and the session can be accepted without exceeding the upload capacity of this cluster. For this, we assume n the average number of videos that a HB can stream in parallel. The average number of supported streams $\mathrm{S_i}$ in a cluster for the video i is:

$$S_i = \frac{1}{C} * P_{|HB|}(i) * |H| * n \qquad (5)$$

The number of the preceding requests for the video i shouldn't then exceed $\mathrm{S_i} - 1$. We can then derive the hit ratio $\mathrm{h_j}$ at the j-th request, as follows:

$$h_j = \begin{cases} \sum_{i=1}^{|V_P|} \frac{P_N(i) * P_{|HB|}(i)}{C} & if \;\; j < S_i \\ \sum_{i=1}^{|V_P|} \left[\frac{P_N(i) * P_{|HB|}(i)}{C} * \sum_{l=0}^{S_i - 1} P_N(i)^l * (1 - P_N(i))^{j-l-1} \right] & otherwise \end{cases} \qquad (6)$$

The hit ratio over the R requests is:

$$h(R) = \frac{1}{R} \sum_{j=1}^{R} h_j \qquad (7)$$

The final expression of hit ratio is then:

$$h(R) = \frac{1}{R}\sum_{j=1}^{R}\sum_{i=1}^{|V_P|}\left[\frac{P_N(i) * P_{|HB|}(i)}{C} * \sum_{l=0}^{S_i-1} P_N(i)^l * (1 - P_N(i))^{j-l-1}\right] \qquad (8)$$

Consider c_{hb} the average cost to access the video from a HB and c_s, the average cost to access the video from the closest CP. The total cost of our video delivery model is then:

$$cost = R * \left(h(R) * c_{hb} + \left(1 - h(R)\right) * c_s\right) \qquad (9)$$

Note that, the cost does not comprise the HBs' caches maintenance. This process is made offline, in the time in which the traffic decreases. Indeed, many works [24] have been carried to study the user behavior in large scale VoD systems and concluded that the user access follows a clear daily pattern. Another important conclusion of this work is that the users' arrival rate varies significantly over the day time. Generally, the access pattern is subject to important peaks during breaks, after work and during the week-end. It is why the maintenance of the HBs caches is scheduled when the video traffic decreases, i.e., in the early morning, according to the study.

4.3.2 Evaluation of the Network- & Context-Aware Anycast-based Server/Peer Selection Mechanism

The performance evaluation of the anycast server selection combined with HB caching has been done under the Network Simulator NS2. The network topology is derived from the Survivable Network Design Library (SNDlib) [26] that count 22 IP backbone networks. We have selected the FRANCE instance. We kept the node placement and the link assignment. However, we made our own assumptions on the service demand and the links capacities have been fixed to 2.5 Gbps. To each node, we have attached 20 HBs with 70 Mbps downlink capacity and 30Mbps uplink capacity. To the HBs, we have attached 1 to 4 clients, with an average of 2 clients per HB. The clients' connectivity varies from 0.5 to 100 Mbps. We have also attached four contents servers, SR and Service Manager nodes to the highest degree routers with 2.5 Gbps connectivity. **Figure 22** illustrates the simulation topology.

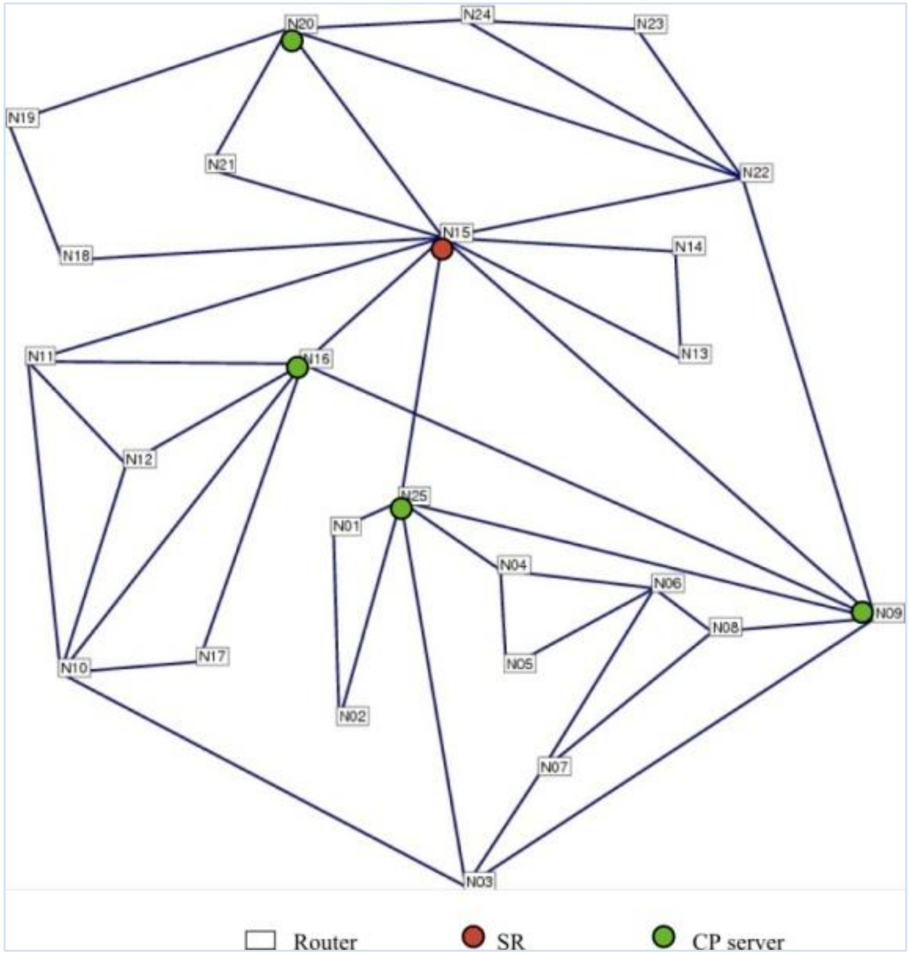


Figure 22: Simulation Topology

Concerning the served video contents, we have considered 500 videos. All videos are present in each server. A video can be streamed in three resolutions: 352x288, 720x576 and 1408x1152 and in three different bitrates for each resolution. The 100 most popular videos are cached within the HB layer. Each HB can hold two videos. The video frequency at the HB layer fits the zipf-like distribution. All the 100 videos are uniformly distributed among the HBs of the overall topology in the order of their popularity.

The video popularity is also zipf-like distribution. Each client requests one video during the entire simulation time. The clients' requests are generated in a poisson model during 250 s which is also the video duration. All the videos have the same duration. For each request, the probability to request the video in the three resolutions presented above is respectively 1/5, 2/5, 2/5. The video bitrate is selected according to the client connectivity.

The proposed delivery approach is evaluated in two scenarios: the first one, with the support of the 50% of HBs and the second, with the support of all the HBs (100% of HBs). To evaluate the approach, we have also compared it to the server selection without any HB support and the widely studied Server Load based selection. All the scenarios are simulated in the same environment. However, since the Server Load selection does not consider any context-aware features, at most two videos can be served in the maximum resolution to the same HB, in order to avoid rapidly overloaded situations. We have compared the four scenarios according to different metrics, namely: the average Server Load, the RTP Path Distance in terms of delay and hops number, and finally, the RTP packet loss.

Figure 23 illustrates the evolution of the average Server Load for the four scenarios through the simulation time. We can easily notice that the more HBs support the video delivery, the more the Server Load is saved. The average server charge is just 50% with 100% HBs support and 60% with 50% HBs support, while it reaches the 90% without HBs support and 85% with the Server Load selection. The system cost is significantly reduced with the HB support.

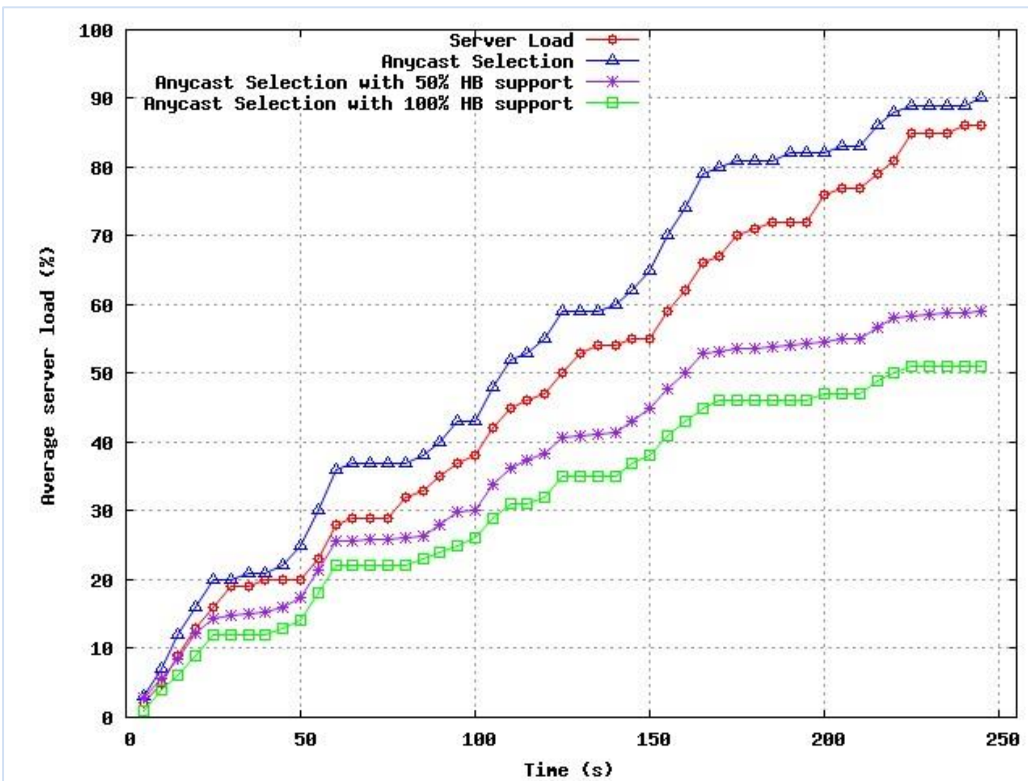


Figure 23: Sim 2 - AverageServer Load vs Simulation Time

Figure 24 clearly shows that our approach avoids losses. It also shows that the higher is the distribution, the less are the probabilities for packet loss to occur. While the packet loss reaches 10% when having all the sessions actives under the Server Load based selection, it is reduced to 5%, 0.2% and 0.15% under our server selection and this with respectively 0%, 50% and 100% HBs support. The packet loss decreasing is the result of avoiding congestion (1) at the user level, by considering the user context, (2) at server level, by considering the Server Load and (3) at network level, by bringing the contents closer to the clients and taking into account the delay metric to select from which server the client will access the content.

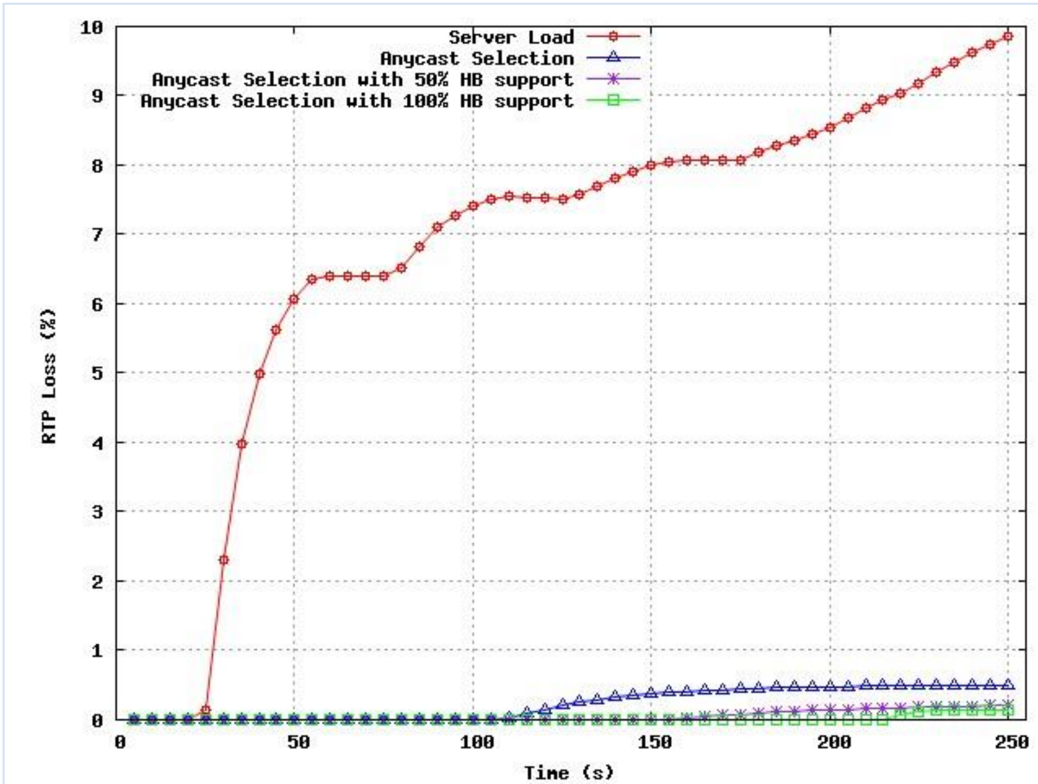


Figure 24: Sim 2 - Average RTP Packet Loss vs Simulation Time

Figure 25 and **Figure 26** illustrate the Path Distance in terms of delay and number of hops metrics. We can notice that the Path Distance is significantly reduced for both metrics under our selection approach. The delay, close to 110 ms under the Server Load based approach, decreases to around 30, 20 and 15 ms under our approach with respectively 0%, 50%, 100% HBs support. This result is due to the delay metric in the metric-based selection phase and to the HBs support for video delivery. The contents are then closer to clients and the delays are accordingly shorter. We can also observe that the evolution of the Path Delay is slow under our approach, while it increases rapidly under the Server Load based selection strategy.

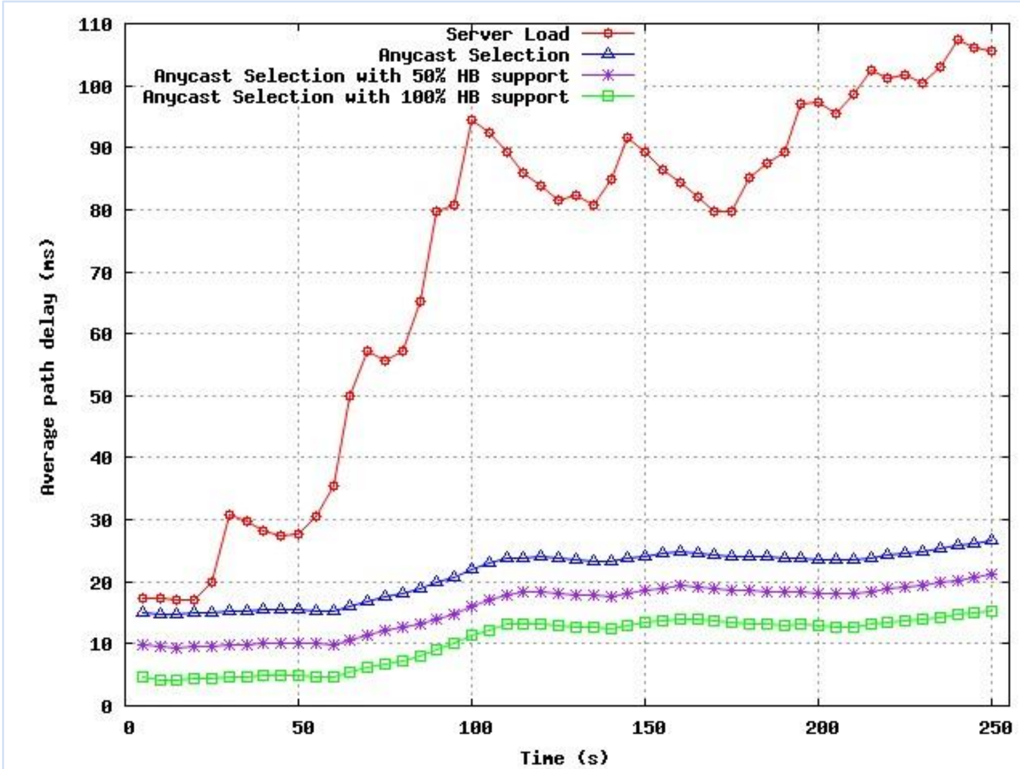


Figure 25: Sim 2 - Average path delay vs simulation time

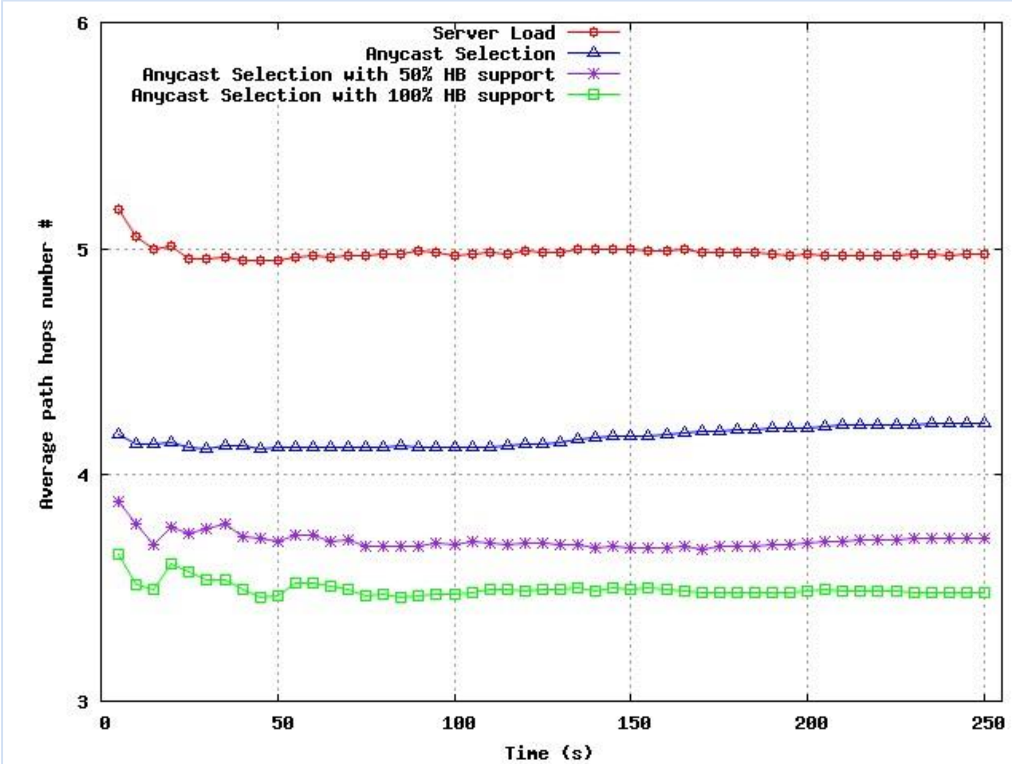


Figure 26: Sim 2 - Average hop number path vs simulation time

4.4 An Online Popularity-based Video Caching Strategy

The strategy proposed to cache videos online inside the Home-Boxes based on popularity derives from and is complementary to the abovementioned work in Chapter 4.3, which considers an offline placement. Now, the objective is to extend capacities of already deployed CDNs, for example, so that they take benefit of the caching and streaming capabilities provided by the Home-Box layer.

4.4.1 The Online Popularity-based Video Caching Strategy

This solution [23] extends the previous one in two ways: 1) by considering a peering CDN that extends the CDN capacity by introducing the Home-Box (HB) equipment in the service delivery chain and 2) the design of an adaptive popularity-based video contents caching strategy. The HBs form then a managed P2P overlay able to support CDN servers in the delivery process, as shown in **Figure 27**. Video contents are dynamically cached in HBs following users' demands, allowing their delivery to End-Users from optimal places. The idea is to leverage the participating and already deployed Home-Boxes caching feature and uploading capabilities to achieve service performance, scalability and reliability, especially in current context where the broadband providers are heavily investing to build out their high speed last mile networks.

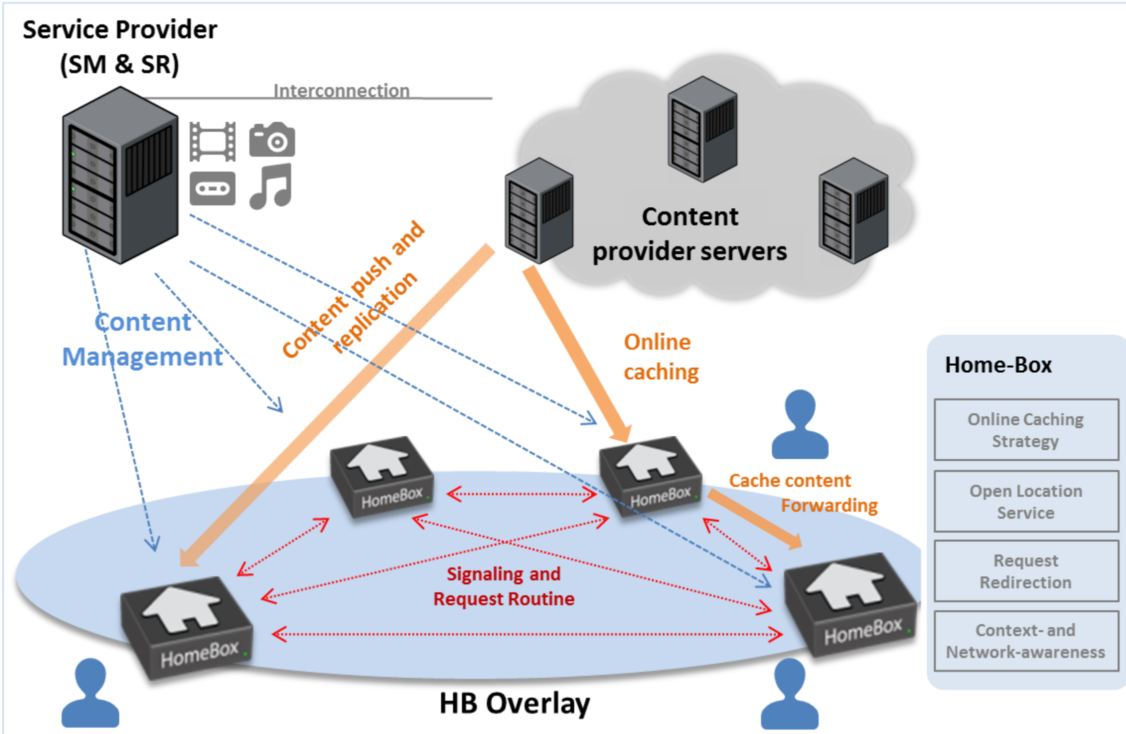


Figure 27: Online Collaborative Caching Approach

Recently, some studies on peer-assisted CDNs have been proposed. The great advantage of such solutions have been demonstrated analytically [29], by simulation [30], or with real implementation and deployment at a large scale [31]. Our proposal differs from these works by relying on the User Environment's storage and connectivity without directly involving clients' nodes participation. The peer node, in our model, is the HB that, despite the fact that it is located in the User's Environment, it still acts as the always-connected "hub" between the User, Network and Service Environments, respectively being capable of shared ownership between the three actors (End-User, Service Provider and Network Provider). Therefore, the HB can be easily managed and monitored by the respective actors, feature that cannot be possible with End-User terminals. Our model is then a fully managed model and, consequently, overcomes the issues induced by the P2P part in this kind of hybrid system (e.g., free riding, reliability, peer churn issues, etc.).

Enabling a distributed edge content hosting is considered as the next step in content distribution paradigm [32]. Recent works have proposed architectures that rely on boxes deployed at the edges of the network, close to users' terminals, for live video streaming services [33][34] or VoD services [35][36]. However, in the latter, video contents are placed offline, which involves an additional delivery cost. In our proposed solution, the videos are placed during their consumptions by End-Users. In addition, an efficient spread of video content copies among the HBs is also considered. The HB clustering and cooperation that we consider in the proposal will also reduce the frequency of the video contents replacement in a single HB cache.

The designed caching strategy aims to make available the video contents to the most possible users' demand, while reducing delays and network bandwidth consumption. Towards this, meeting the users' demand is one of the main requirements of any replication strategy; therefore, the video popularity is one of the most important parameter considered in our design. The more popular is the video, the bigger is its frequency in the HB overlay. As explained previously, the video popularity is periodically tracked by the Service Manager. For our analysis, the video popularity fits zipf-like distribution. Considering M videos, ranked in the order of their popularity from 1 to M, video 1 being the most popular, the probability $\mathrm{V_i}$ to request the video of rank i is:

$$\boldsymbol{P_M(i) = \frac{1/_{i^\alpha}}{\sum_{k=1}^{M} 1/_{k^\alpha}}} \qquad (\mathbf{10})$$

where, α represents the skew of the distribution.

The number of HBs that will hold each video in each cluster is then directly related to the requesting probability of this video.

Assuming N, the number of participating HBs to the overlay, $\mathrm{M_p}$, the number of the most popular videos considered in the caching strategy, the number $\mathrm{N_i}$ of HBs that will cache the video $\mathrm{V_i}$ is:

$$\boldsymbol{N_i = ceil\left(\frac{P_M(i)}{\sum_{j=1}^{M_p} P_M(j)} * N * C\right)} \qquad (\mathbf{11})$$

where, $\mathrm{M_p}$ is assumed large enough so that $\mathrm{N_i}$ never exceeds N.

It is important to replicate videos in the overlay according to their popularity; however, another important issue is how to better distribute them among the HBs. Towards an efficient distribution, we have used the K-means algorithm to determine the distance to respect between any two HBs that will cache the same video content. Given a set of points P, the k-means clustering problem seeks to find a set K of $\mathrm{k}$centers, so that $\sum_{\mathrm{p \in P}} \mathrm{d(p, K)^2}$ is minimized. The k-means problem is NP-hard. However, there exist approximation algorithms that can be used.

In our case, HBs sub-clusters are calculated for each video following the order of their popularity. If we consider the video V_i, the parameters P and K of the K-means algorithm are represented respectively by the set of HBs that have enough place in their caches and by the N_i HBs that should hold it. Once the N_i sub-clusters for V_i are formed, the maximum Network Distance between the couples of nodes under each sub-cluster c_i is calculated and the average of these Network Distances d_i is assigned to V_i. d_i represents the Network Distance that will separate any couple of HBs that will cache the video V_i. At the end of this process, a set of Network Distances $\{d_1, d_2, \dots, d_{M_p}\}$ are then assigned to the set of the M_p most popular videos $\{V_1, V_2, \dots, V_{M_p}\}$. Since the HBs are considered stable components, this process should not happen frequently. It can be performed offline when the percentage of HBs that have joined or left the network exceeds a certain threshold, identified as a potential threat for system performance.

In addition to the video caching frequency, another major design decision is to determine where and when the video has to be either cached or removed in order to be replaced by another one. Some works [37][38] considered that the cache replacement issue was less important and the main presented argument is that the caches are large enough to host most of the requested contents. However, with the explosive growth experienced by multimedia contents, that are known to require large storage and bandwidth resources, this argument cannot remain valid and the cache size becomes again a limit to build lucrative video streaming services. This limit is all the more important with the advent of User Generated Contents (UGCs) that bring an unlimited choice of videos to the Service Providers libraries, since enabling millions of users to be not only content consumers but also producers.

Before explaining our proposal related to video placement and replacement issue, we would first present briefly an overview of cache replacement techniques that complete the caching techniques already presented in this chapter.

Taxonomy on Web cache replacement strategies is given in [39]. The latter presents the following classes:

- Recency-based strategies replace, from the cache, the objects that have not been requested for the longest time. Most of them are extensions of the well-known Least Recently Used (LRU) strategy such as LRU-Min [40], PSS [41], LRU-LSC [42], etc. The latter exploits the temporal locality of users' requests as a critical factor for objects purge from the cache ;
- Randomized strategies are the simplest approach to implement. The aim is to reduce the complexity of replacement strategies without sacrifying quality too much ;

- Frequency based strategies are based on the frequency replacement factor. These strategies are direct extensions of the Least Frequently Used (LFU) strategy. They exploit the fact that objects have different popularities that are reflected on the frequency of users' requests for these objects. The objects popularity can be maintained for either all objects or only the objects that are present in the cache ;
- Recency/Frequecy-based strategies are more complex strategies that mix recency and frequency to designate the object that will be removed from the cache. Example of such strategies are SLRU [43], LRU* [44], LRU-Hot [45], etc. ;
- Function-based strategies evaluate a specific function that combines different factors such as recency, frequency, object size, cost of fetching the object, etc. Most of these factors are weighted. The weights are commonly deduced from usage traces.

In this proposal, we aim to design an adaptive and cooperative popularity-based video placement and replacement strategy. The decision to place or replace the video in the cache of the requesting HB depends not only on the videos which are present in the cache of that HB but also on the contents cached inside its neighbor HBs. The set of neighbor HBs is determined according to the Network Distances estimation, as presented in the previous section.

In order to make the model tractable, we have made the following assumptions:

- All HBs have the same storage capacity and upload connectivity, respectively represented by C and n (n is expressed in term of average simultaneous supported streams);
- We assume stable the HBs connectivity during all the requests for which we formulate the HB overlay hit ratio (percentage of requests satisfied by HBs);
- The users' requests are independent and fit a poisson model with a mean rate of λ requests per time unit;
- The videos length is represented by the mean duration of T time units;
- We consider that the HB overlay is already organized in L geographical clusters in which the proposed video replication solution is considered separately.

Considering then the j-th request for service/content consumption, the HB caches a video $\mathrm{V_i}$ if the following conditions are satisfied:

- The HB is requesting the video V_i, which is estimated by the probability $P_M(i)$;
- The closest HB that caches V_i is located at a Network Distance greater than d_i from it, which means that, at request j, V_i has not yet been cached in the sub-cluster c_i of the requesting HB at a previous request to j.

The HB has either sufficient available cache capacity to store $\mathrm{V_i}$ or at least one of the C cached videos at its side is less popular than $\mathrm{V_i}$. In this case, the latter will be replaced by $\mathrm{V_i}$.

More formally, assuming $\mathrm{Pb_c(i,j)}$ is the mean probability that $\mathrm{V_i}$ had been cached at a previous request m to j and had not been replaced till j, $\mathrm{Pb_c(i,j)}$ can then be derived as follows:

- In both the initial phase where the HB cache is not full ($j \le C$) and the case where V_i is one of the C most popular videos ($i \le C$), V_i is cached if the following conditions are satisfied:

  $\mathrm{V_i}$ is requested at m (which is estimated by $\mathrm{P_M(i)}$) ;

  and $\mathrm{V_i}$ has not been already cached in the cluster $\mathrm{c_i}$ of the requesting HB at a previous request to m (which is estimated by $\frac{1}{\mathrm{N_i}} * \left(1 - \mathrm{P_c(i,m)}\right)$);

- Otherwise, we derive first the probability that either $\mathrm{V_i}$ hasn't been cached at any previous request to j or $\mathrm{V_i}$ has been cached at a previous request m to j but has been replaced before j. In the first case, $\mathrm{V_i}$ either has not been requested $\left(1 - \mathrm{P_N(i)}\right)$ or it has been requested but has already been cached in one of the HBs that belongs to the cluster $\mathrm{c_i}$ of the requesting HB $\left(\frac{\mathrm{P_c(i,m)}}{\mathrm{N_i}}\right)$. As well, the cache of the HB might also be full, meaning that the cache holds C more popular videos than $\mathrm{V_i}$ before the request m. This is estimated by: $\left(\frac{1}{\mathrm{N}} * \frac{1}{\mathrm{i-1}} \sum_{\mathrm{l=1}}^{\mathrm{i-1}} \mathrm{P_c(l,m)}\right)^{\mathrm{C}}$. In the second case, where $\mathrm{V_i}$ is cached but replaced before j, the HB at request j should have in its cache C more popular videos than $\mathrm{V_i}$, which is estimated by $\left(\frac{1}{\mathrm{N}} * \frac{1}{\mathrm{i-1}} \sum_{\mathrm{l=1}}^{\mathrm{i-1}} \mathrm{P_c(l,j)}\right)^{\mathrm{C}}$.

To summarize, $\mathrm{Pb_c(i,j)}$ is then:

$$
P_c(i,j) = \begin{cases}
\dfrac{1}{j-1} \displaystyle\sum_{m=1}^{j-1} P_M(i) * \dfrac{1}{N_i}\left(1 - P_c(i,m)\right) & if\left((i \le C) || (j \le C)\right) \\
\dfrac{1}{j-1} \displaystyle\sum_{m=1}^{j-1} \left(1 - \left(\left(\begin{array}{c} \left(1 - P_M(i)\right) + \\ P_M(i) * \left(\begin{array}{c} \left(\dfrac{P_c(i,m)}{N_i} + \left(\dfrac{1}{N} * \dfrac{1}{i-1} \displaystyle\sum_{l=1}^{i-1} P_c(l,m)\right)^C \right) + \\ \left(\left(1 - \left(\dfrac{P_c(i,m)}{N_i} + \left(\dfrac{1}{N} * \dfrac{1}{i-1} \displaystyle\sum_{l=1}^{i-1} P_c(l,m)\right)^C \right) \right) * \\ \left(\dfrac{1}{N} * \dfrac{1}{i-1} \displaystyle\sum_{l=1}^{i-1} P_c(l,j)\right)^C \right) \end{array} \right) \end{array} \right) \right) \right) & otherwise
\end{cases} \quad (12)
$$

The probability that the video $\mathrm{V_i}$ is present in k HBs at request j is then:

$$P_c(i,j,k) = \frac{C_{j-1}^{k} * P_c(i,j)^k * \left(1 - P_c(i,j)\right)^{j-k-1}}{\sum_{l=0}^{N_i} C_{j-1}^{l} * P_c(i,j)^l * \left(1 - P_c(i,j)\right)^{j-l-1}} \qquad (13)$$

In as much as $\mathrm{N_i} * \mathrm{L}$ is the maximum number of HBs that can hold $\mathrm{V_i}$ in the whole HB overlay.

The efficiency of a replication strategy is a trade-off between the service scalability and quality, from one side, and the operating cost of the system, from the other side. To be efficient, our strategy thus needs to satisfy the maximum simultaneous users' requests by the HB overlay, while keeping quality at satisfying threshold for users. Indeed, making the videos reachable from the HB overlay reduces considerably the path delay, saves network and server bandwidth and consequently reduces packet loss. The goal is then to maximize the HB overlay hit ratio. Toward this goal, and considering, on one hand, the high number of video contents that constitute the VoD catalogues (especially with the advent of User Generated Contents UGC), and, on the other hand, their extremely variant popularities, it is important to model the relationship between the number of the most popular videos that will be cached in the HB overlay and the system cost which is directly related to the HB overlay hit ratio. The way to formulate such equation thus constitutes a key modelling issue.

To this end, we have derived the HB overlay hit ratio for the case of finite HB overlay capacity N and finite number R of requests. Considering the hit ratio $\mathrm{h(HB, j)}$ of the j-th request (where $1 \le \mathrm{j} \le \mathrm{R}$), the request is satisfied by the HB overlay if and only if: 1) the requested video $\mathrm{V_i}$ is present at the HB overlay and 2) the streaming capacity of the cluster for $\mathrm{V_i}$ has not been exceeded. Assume $\mathrm{S_i} = \mathrm{N_i} * \mathrm{n}$ the maximum parallel streams supported by the HB overlay for $\mathrm{V_i}$, λ_i the average number of requests for $\mathrm{V_i}$ per T time units in the cluster with $\lambda_\mathrm{i} = \mathrm{P_N(i)} * \lambda * \mathrm{T/L}$and $\mathrm{M_p}$ the number of the most popular videos that will be cached in the HB overlay, the hit ratio $\mathrm{h(HB, j)}$ of j -th request can be derived as follows:

- Assume m the number of requests for V_i in the last T time units, l the number of HBs that hold V_i , the HB overlay is able to serve another request for V_i if the following conditions are satisfied:
  - l is large enough to serve the m previous request for V_i which means that $l \ge ceil\left(\frac{m+1}{n}\right)$;
  - and because a HB can cache multiple videos, these l HBs must have served at most $(l * n - 1)$ videos V_j that they have already cached (with $\left(j \le M_p\right)$) .

Note that the two previous conditions are for the case that m does not exceed the maximum number of streams that can be supported by the HB overlay. For the case where m exceeds this value, the request cannot be satisfied by the HB overlay since it would be saturated for $\mathrm{V_i}$.

Then, in the particular case where j is large enough that $\forall \mathrm{i} : \mathrm{S_i} < j$ (note that if $\mathrm{S_i} < j$ then $\mathrm{N_i} < j$), the hit ratio $\mathrm{h(HB, j)}$ is derived as follows:

$$\mathrm{h(HB,j)} = \sum_{\mathrm{V_i}} \left(\left(\left(\mathrm{P_M(i)} * \left(1 - \left(\sum_{\mathrm{m=S_i}}^{j-1} \mathrm{e}^{\lambda_i} \frac{{\lambda_i}^m}{\mathrm{m!}} + \sum_{\mathrm{m=0}}^{\mathrm{S_i}-1} \mathrm{e}^{\lambda_i} \frac{{\lambda_i}^m}{\mathrm{m!}} * \left(\sum_{\mathrm{m=0}}^{\mathrm{ceil}\left(\frac{\mathrm{m+1}}{\mathrm{n}}\right)-1} \mathrm{P(i,j,l)} + \left(\sum_{\mathrm{l=ceil}\left(\frac{\mathrm{m+1}}{\mathrm{n}}\right)}^{\mathrm{N_i*L}} \mathrm{P(i,j,l)} * \left(\sum_{\mathrm{s=l*n}}^{\mathrm{j-1}} \left(\mathrm{e}^{\lambda} \frac{\lambda^s}{\mathrm{s!}} * \left(\frac{\mathrm{l}}{\mathrm{N*s}} \sum_{\mathrm{t=j-s}}^{\mathrm{j-1}} \mathrm{h(HB,t)} \right)^{\mathrm{l*n}} \right) \right) \right) \right) \right) \right) \right) \right) \right) \tag{14}$$

More generally, $\mathrm{h(HB, j)}$can be derived for the three cases where $\mathrm{V_i}$ can respectively be part of the three complementary following sets of videos at requestj:

- MP1: if $j < N_i'$: the number of requests received by the system is less than the number N_i' of maximum HBs that can hold V_i in the system (number of caches), then the maximum number of caches are $j - 1$;
- MP2: if $N_i' < j < S_i'$: the number of requests received by the system is greater than the number V_i of HBs that can hold V_i but less than the number S_i' of the parallel streams supported by the HB overlay for V_i, then the maximum parallel active streams supported by the HB overlay for V_i is $j - 1$;
- MP3: if $S_i' > j$: the number of requests received by the system is greater than the number S_i' of the parallel streams supported by the system for V_i which is the particular case described previously.

$\mathrm{h(HB, j)}$ is then the sum of the three previous cased since the request are independent and the cases are complementary :

$$
\begin{aligned}
h(HB,j) = & \sum_{V_i \in MP1} \left(\left(\sum_{m=0}^{j-1} e^{\lambda_i} \frac{{\lambda_i}^m}{m!} * \left(P_M(i) * 1 - \sum_{m=0}^{ceil\left(\frac{m+1}{n}\right)-1} P(i,j,l) + \left(\sum_{l=ceil\left(\frac{m+1}{n}\right)}^{j-1} P(i,j,l) * \left(\sum_{s=l*n}^{j-1} \left(e^{\lambda} \frac{\lambda^s}{s!} * \left(\frac{1}{N*s} \sum_{t=j-s}^{j-1} h(HB,t) \right)^{l*n} \right) \right) \right) \right) \right) \right) \\
& + \sum_{V_i \in MP2} \left(\left(\left(\sum_{m=0}^{j-1} e^{\lambda_i} \frac{{\lambda_i}^m}{m!} * \left(P_M(i) * 1 - \sum_{m=S_i}^{j-1} e^{\lambda_i} \frac{{\lambda_i}^m}{m!} + \sum_{m=0}^{ceil\left(\frac{m+1}{n}\right)-1} P(i,j,l) + \left(\sum_{l=ceil\left(\frac{m+1}{n}\right)}^{N_i} P(i,j,l) * \left(\sum_{s=l*n}^{j-1} \left(e^{\lambda} \frac{\lambda^s}{s!} * \left(\frac{1}{N*s} \sum_{t=j-s}^{j-1} h(HB,t) \right)^{l*n} \right) \right) \right) \right) \right) \right) \right) \\
& + \sum_{V_i \in MP3} \left(\left(\left(\sum_{m=0}^{S_i-1} e^{\lambda_i} \frac{{\lambda_i}^m}{m!} * \left(P_M(i) * 1 - \sum_{m=S_i}^{j-1} e^{\lambda_i} \frac{{\lambda_i}^m}{m!} + \sum_{m=0}^{ceil\left(\frac{m+1}{n}\right)-1} P(i,j,l) + \left(\sum_{l=ceil\left(\frac{m+1}{n}\right)}^{N_i} P(i,j,l) * \left(\sum_{s=l*n}^{j-1} \left(e^{\lambda} \frac{\lambda^s}{s!} * \left(\frac{1}{N*s} \sum_{t=j-s}^{j-1} h(HB,t) \right)^{l*n} \right) \right) \right) \right) \right) \right) \right)
\end{aligned}
\tag{15}
$$

The average hit ratio over the R requests is:

$$h(HB,R) = \frac{1}{R} \sum_{j=1}^{R} h(HB,j) \qquad (16)$$

Now that the hit ratio for the HB overlay is completely derived, let us consider the hit ratio of the CDN server. Since video content are large, we assume the CDN server has a limited capacity of C_S videos with C_S considerably larger than C. When the video cannot be streamed by the HB overlay, it is accessed from a CDN server, if it is part of the C_S most popular videos and it was accessed at least once in the previous requests. The hit ratio of a CDN server $h(S,j)$ is then:

$$h(S,j) = \sum_{i=1}^{C_S} P_M(i) * \left(1 - \left(1 - P_M(i)\right)^{j-1}\right) \qquad (17)$$

We have defined the hits ratio of both the HB overlay and the CDN server. Now let's consider the total cost of the system. Assuming the mean cost to access the video from a HB, a CDN server and a CP being respectively $\mathrm{ct_{HB}}$ $\mathrm{ct_{CDN}}$ and $\mathrm{ct_{CP}}$, the total system cost of the system for the R requests is then:

$$TCost = R * \Big(h(HB,R) * ct_{HB} + \big(1 - h(HB,R)\big) * \big(h(S,R) * ct_{CDN} + \big(1\_h(S,R)\big)ct_{CP}\big)\Big) \quad (18)$$

$\mathrm{ct_{HB}}$ is the lowest of the three costs. Therefore, the total system cost is maximized when the HB overlay hit ratio is maximized. Since we are particularly interesting on the relationship between the system cost and the number of the most popular videos $\mathrm{M_P}$ that will be cached in the HB overlay, we gradually increase $\mathrm{M_P}$ starting by the initial value of $\mathrm{N}/_{\mathrm{C}}$. For each value, we calculate the total system cost TCost. The value $\mathrm{M_P}$ that optimizes the total system cost for the R requests is obtained when $\mathrm{TCost}(\mathrm{M_p}) < TCost(\mathrm{M_p} + 1)$.

4.4.2 Evaluation of the Online Popularity-based Video Caching Strategy

The performance evaluation of the proposed on-line popularity-based video content replication combined with the two-level selection strategy, within the hybrid CDN-P2P content delivery platform has been done under NS2. The hierarchical network topology in which we have simulated our cache strategy was generated with Brite in the Waxman model. The generated topology is fully conforming to the architecture presented. It consists of 200 routers with constant connectivity of 1Gb/s. The routers are placed in 5 ASs. In average, 5 HBs are attached to each router with a connectivity of 100 Mb/s. To simulate a real home network system, 1 to 4 clients are attached to each HB with a connectivity that varies from 0.5 to 100 Mb/s. We have placed, in each AS, a CDN server with a 2Gb/s connectivity and attached it to the router with the highest degree (highest number of routers linked to it). The topology also comprises Service Registry and Service Manager nodes that are also linked to the routers with the highest degree.

The video catalogue is composed of 10000 videos provided in three different formats:

- CIF for mobile terminals (352x288);
- 4CIF to provide TV-like quality (720 x576);
- 16CIF to provide High Definition video contents (1408x1152).

The probability to request the video in each format is respectively 1/5, 2/5 and 2/5. The video popularity fits the zipf distribution with a skew of 7.33.

Only the 1000 most popular videos are cached in the HB overlay. The others are only provided by CDN servers.

10000 requests are generated in a poison model during 30 min with an average of 400 requests per min, which implies situations of 2000 parallel streams. The video duration is set to 5 min. The clients linked to each HB request the VoD system 4 to 16 times with an average of 10 requests per HB.

All the HBs participate to the overlay and each of them can cache 5 videos and can provide in average 5 parallel streams. At the beginning of the simulation, all the caches of the HBs are empty.

In order to evaluate the efficiency of the proposed solution, it is also compared to a pure CDNs platform composed only with CDN servers. The two scenarios are simulated in the same conditions with the same traffic model. The two approaches are compared according to the following metrics: the average server load and the data path delay.

The overlay hit ratio, presented in **Figure 28**, represents the ratio of requests that are satisfied by HBs. As illustrated by the figure, in the initial phase of the simulation, the hit ratio is low. The reason is that the videos were not yet placed in the HBs caches. In our cache strategy, the videos are progressively placed in the HBs' caches, while they are requested and consumed by End-Users. After almost three video durations (~900s), we can note that the hit ratio increases continuously until the end of the simulation. Thus, the more the HBs' caches get filled with videos, the highest gets the HB overlay hit ratio.

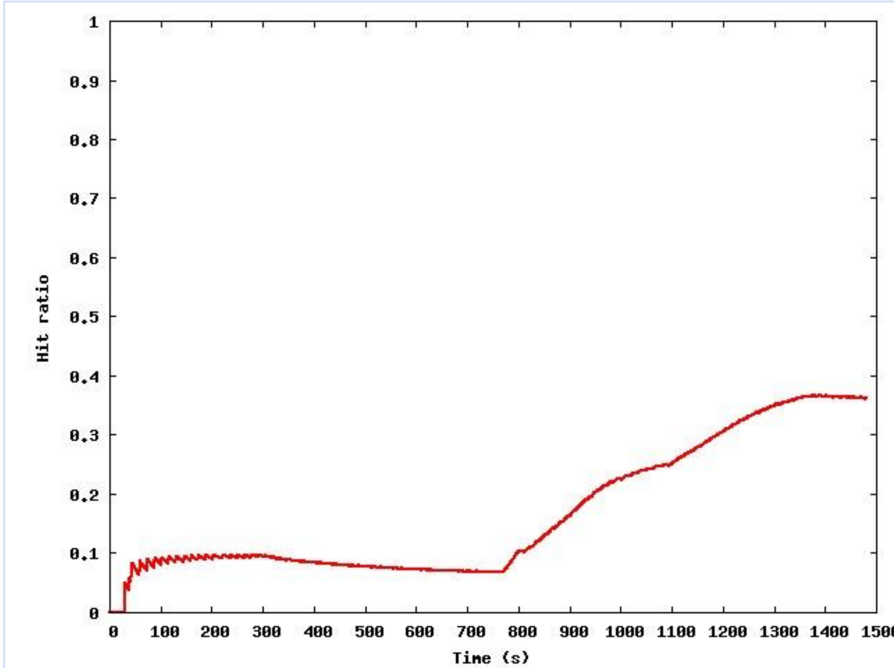


Figure 28: HB Overlay Hit Ratio

Figure 29 represents the average loads of CDN servers. We can note that our approach saves approximately 70% of the Servers' Load at the end of the simulation. In the initial phase of the simulation, the two Server Loads are almost the same while as the HB caches get filled with videos, the difference between them increases. This result is in line with the increase of the HB overlay hit ratio and permits to show the benefit of our solution towards CDN Servers' Load, high impacting metric in investment strategies for SPs.

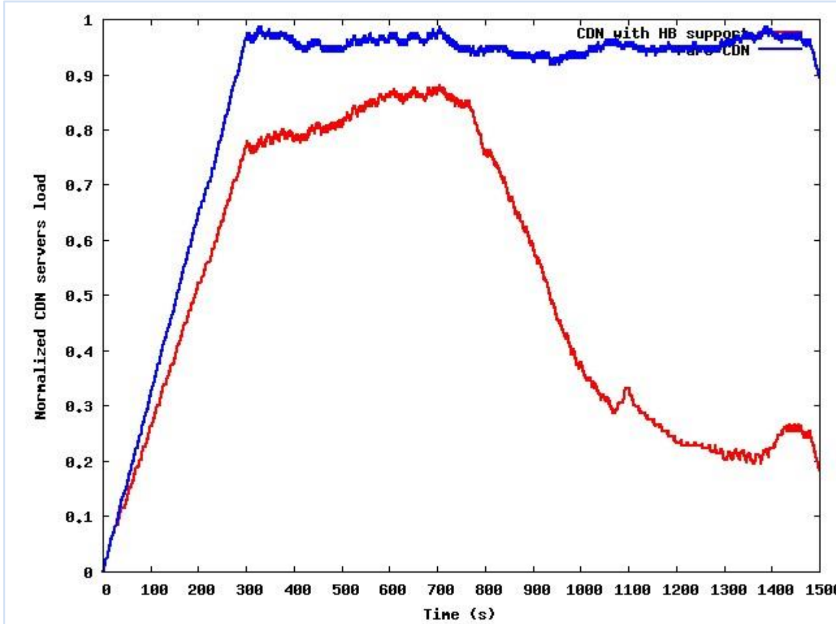


Figure 29: Sim3 - Average CDN Servers Consumed Bandwidth

Another system cost metric considered in the evaluation of our strategy is the Path Delay. This metric represents the average delay of RTP data packets of all the engaged video sessions in the system at a given moment. As expected and illustrated in **Figure 30**, the Path Delay is always lower under our approach. However, the difference becomes even more significant as the simulation time increases and the videos are more and more distributed from the HBs' caches. This result is in line with the increase of the HB overlay hit ratio and permits to show the benefit of our solution towards the expected quality of the video service, high impacting metric in service adoption by End-Users.

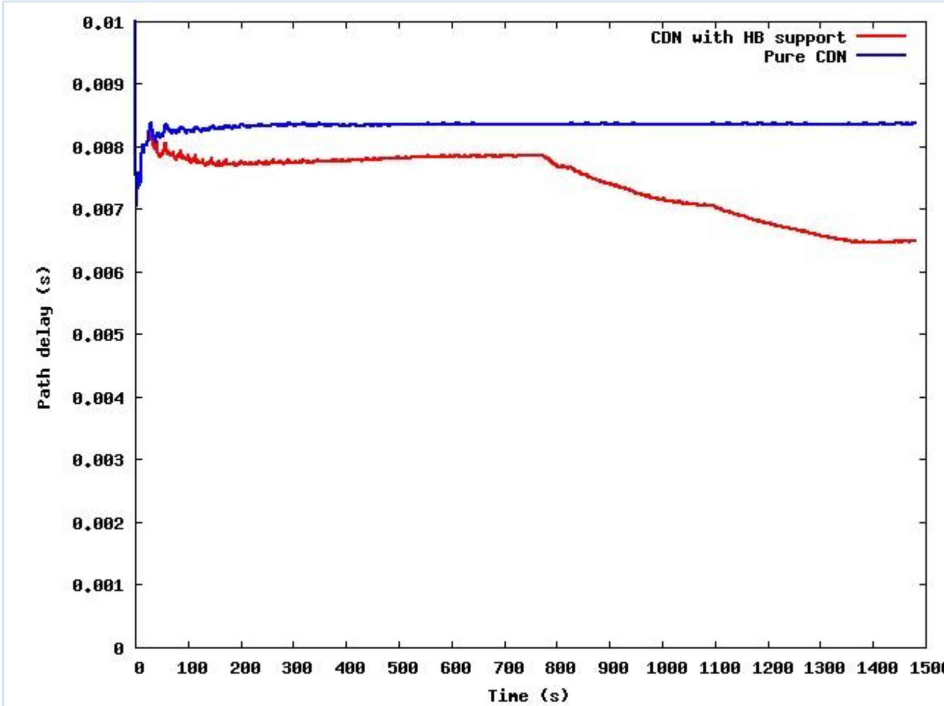


Figure 30: Sim3 - Average Data Path Delay

We can also mention that our approach does not need an additional bandwidth for caching the served videos in the HBs' caches, since the caching process is performed while consuming the videos. Contrary to offline video placement approaches that require additional bandwidth for the placement of video contents in caches, the total bandwidth cost of our solution is exactly the necessary bandwidth for the service/content consumption.

To conclude, the proposed context-aware video delivery solution to be used within the proposed Future Media Internet architecture presents major innovations on the design of (1) an enhanced service oriented architecture based on a virtual Home-Box overlay able to assist the video delivery infrastructure. The idea is to leverage the User Environment capabilities, as well as the Service/Content Provider features, in order to reach high-efficiency in delivering large-scale and high quality VoD services. The foreseen solution benefits from the self-scaling property of the P2P systems, while keeping the control and quality guarantees of managed infrastructures; (2) a popularity-based video replication strategy over the HB overlay that allows an efficient distribution of the video contents among the HBs cache as well as an adaptive and cooperative video content replacement strategy; (3) an efficient server selection strategy that copes with the video services requirements and the End-User experience. Indeed, the proposed strategy combines multiple filters based on both context-aware policies and accurately measured metrics for selecting the most suitable server for each client request.

The evaluation of our proposal through simulation shows the gain of bandwidth, storage and service quality induced by such solution. As well, the obtained results clearly highlight the promising aspects of considering the Home-Box node in the video delivery chain.

Another advantage of the proposed video delivery scheme is its ability to be easily integrated in the next generation networks platforms since the HB is a component stemming from already deployed home devices (home gateways, set-top boxes).

One issue in the deployment of the proposed infrastructure and associated solution relies in its strong dependency to the last mile network resources. However, the progress deployment of Fiber-To-The-Home (FTTH) and the investments made by ISPs in the deployment of more powerful home gateways and devices encourage such a solution.

4.5 Conclusions on the Home-Box-assisted Content Delivery Network System

There are several reasons and advantages to incorporate a media caching mechanism into the Home-Box entity. Those reasons/advantages arise from the limitations of current content distribution platforms, namely (1) CDNs, which represent a high cost for video content distribution; (2) Traditional P2P, typically unreliable, with lack of control and management.

Those limitations are particularly relevant in the case of User-Generated Content (UGC) distribution and High-Popularity VoD use cases. The former is true since the Service Providers are not keen to invest heavily in network infrastructures and back-office centres to distribute content that normally does not generate much revenue. In that scenario, the simple employment of a CDN presents prohibitively high costs and is not seen by the Service Provider as the solution for this problem.

In the case of High-Popularity VoD assets distribution, the distributed caching at the edge nodes is seen as a very welcome add-on to the traditional cluster of VoD servers. It is foreseen that this complement to the normal cluster of VoD server would burst the capability of the system in terms of media distribution (would be able to handle more easily the spikes of requests) reducing, at the same time, the amount of OPEX and investment.

Therefore, the proposal of an efficient caching and replication strategy within Home-Boxes for assisting and enhancing (by considering user context- and network-related information) Content Delivery stands as an evolutionary solution for Media distribution stakeholders. **Table 10** summarizes the main differences between the proposed solution and the traditional CDN approach.

Table 10: Comparison between CDN and HB-assisted CDN

Function	CDN	HB-assisted Solution
Push/Pull-based Replica provisioning	Mainly push-based	Both
Replica provisioning algorithm	Access pattern Network topology Latency-optimized	+ Popularity-based frequency + HB distance
Cooperative/uncooperative Request redirection	Cooperative	Cooperative + Adaptive delivery + Multi-source delivery
Cooperative/uncooperative CachingStrategy	Uncooperative Central management or Local replacement	Cooperative Collaborative online caching

The proposed **Home-Box Assisted CDN System** with its two innovative features for server selection and online caching has been designed, evaluated analytically and through simulations and has produced interesting results for consideration by the relevant stakeholders, as described in the sections above. In addition, a real implementation has been done as well as extensive trials and evaluations in real conditions, all through the European Project ALICANTE. All the related information can be found in [2][46][47].

Considering the network-aware peer selection process, further work has been performed relying on the cooperation with the CAN layer at the monitoring level, in order to establish a more accurate **"Network Distance Cost"** evaluation and to rely on it for the **"Home-Box Distance"** value (between two Home-Boxes peers). This work and its results can be found in [48] [49] [2][47].

Additional work has also been performed within the Home-Box element in order to enable adaptation features, during the delivery phase, while the content is being consumed. This process can go from layer dropping up to transcoding or server switching. It is detailed in Annex 2 and Annex 3, [50]. Towards complete adaptation, serious performance bottlenecks have been raised and work has been conducted within [51] and [52] to go and investigate a software/hardware co-design approach, with FPGA involvement and dynamic reconfiguration capabilities, inside a Home-Box.

Bibliography

[1] S. Ait Chellouche, "Context-aware Media Services delivery in heterogeneous environments for Future Media Networks", PhD Thesis, http://ori-oai.u-bordeaux1.fr/pdf/2011/AIT_CHELLOUCHE_SORAYA_2011.pdf, 2011.

[2] M. Sidibé, D. Négru, (eds.) et al., ICT ALICANTE, Deliverable D4F, "The ALICANTE Home-Box Layer – Final", http://www.ict-alicante.eu/validation/download/work-package/alicante_d4f_v1.0.pdf, September 2013.

[3] J. Wang, "A survey of web caching schemes for the Internet", ACM SIGCOMM Computer Communication Review Volume 29 Issue 5, October 1999.

[4] B. D. Davison, "A Web Caching Primer", IEEE Internet Computing Journal, Volume 5 Issue 4, July 2001

[5] D. Zeng, F. Y. Wang, M. Liu, "Efficient web content delivery using proxy caching techniques", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, volume 34 issue 3, pp. 270-280, 2004

[6] L. Kagal, T. Finin, A. Joshi, "A Policy Based Approach to Security for the Semantic Web", In 2nd International Semantic Web Conference (ISWC2003), September 2003

[7] C. Partridge, T. Mendez and W. Milliken, "Host Anycasting Service", RFC1546, 1993

[8] T. Hardie, "Distributing Authoritative Name Servers via Shared Unicast Addresses", RFC 3258, April 2002

[9] C. Huitema, "An Anycast Prefix for 6to4 Relay Routers", RFC3068, June 2001

[10] D. Katabi, J. Wroclawski, "A Framework for Scalable Global IP-Anycast (GIA)", SIGCOMM'00, Stockholm, Sweden, 2000

[11] W. Xiaonan, Q. Huanyan, "An Anycast Communication Model on Solving Scalability in IPv6", Control, Automation, Robotics and Vision, 2006.

[12] H. Miura, M. Yamamoto, "Server Selection Policy in Active Anycast", IEICE Trans. Commun., Vol. E84.B, No. 10 October 2001

[13] H.B. Hashim, J.A. Manan, "An active anycast rtt-based server selection technique", Proceedings of 13th IEEE International Conference on Networks 1 145–149, 2005

[14] H. Ballani, P. Francis, "Towards a global IP Anycast service", Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications 301–312, 2005

[15] T. Stevens, J. Vermeir, M. De Leenheer, C. Develder, F. De Turck, B. Dhoedt, P. Demeester, "Distributed Service Provisioning Using Stateful Anycast Communications", lcn, pp.165-174, 32nd IEEE Conference on Local Computer Networks, 2007

[16] S. Bhattacharjee, M.H. Ammar, E.W. Zegura, V. Shah, and Z. Fei, "Application-Layer Anycasting," IEEE INFOCOM'97, Kebe, Japan, April 1997

[17] E. Zegura, M. Ammar, Z. Fei, and S. Bhattacharjee. "Application-layer anycasting: A server selection architecture and use in a replicated web service". ACM/IEEE Transactions on Networking, 8(4):455-466, 2000

[18] Z. Fei, S. Bhattacharjee, E.W. Zegura, M. Ammar, "A Novel Server Selection Technique for Improving the Response Time of a Replicated Server", INFOCOM, 1998

[19] Zheng da Wu, Chris Noble, and D. Huang. "Optimal Video Distribution Using Anycasting Service", Information Technology papers, 1999

[20] H.-Y. Kung, C.-M. Huang, H.-H. Ku, C.-Y. Lin, "Efficient multimedia distribution architecture using anycast", Computer Communications, v.30 n.2, p.288-301, Jan. 2007.

[21] S. Ait Chellouche, D. Négru, E. Borcoci, E. LeBARS, "Anycast-based Context-Aware Server Selection Strategy for VoD Services", in Proc. of IEEE GLOBECOM Workshops (GC Wkshps), Miami, Florida, 2010

[22] S. Ait Chellouche, D. Négru, E. Borcoci, E. LeBARS, "Context-Aware Distributed Multimedia Provisioning based on Anycast Model towards Future Media Internet", in Proc. of 8th IEEE Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, 2011

[23] S. Ait Chellouche, D. Négru, Y. Chen, "Home-Box-assisted Content Delivery Network for Internet Video-on-Demand Services", in proc. of the IEEE Symposium on Computers and Communications (ISCC'12), Cappadocia, Turkey, July 2012.

[24] H. Yu, D. Zheng, B. Zhao, W. Zheng, "Understanding User Behavior in Large-Scale Video-on-Demand Systems," Proc. EuroSys Conf., 2006

[25] C. Griwodz, M. Bar, L.C. Wolf, "Long-term Movie Popularity Models in Video-on-Demand Systems or, the Life of an On-Demand Movie," Proc. Multimedia 97, ACM Press, New York, pp. 349-357, 1997

[26] S. Orlowski, M. Pióro, A. Tomaszewski, R. Wessäly, "SNDlib 1.0-Survivable Network Design Library", Networks 55:3, pp. 276-286, 2010

[27] S. Ait Chellouche, D. Négru, Y. Chen, W. Aubry, "Home-Boxes Support for an

Efficient Video-on-Demand Distribution", in proc. of the 2011 IEEE International Conference on Multimedia and Expo (ICME 2011), WoMAN Workshop, Barcelona, Spain, July 2011.

[28] S. Ait Chellouche, D. Négru, Y. Chen, "Home-Box-assisted Content Distribution Network for Internet Video-on-Demand Services", Technical report, 2012

[29] D. Xu, S. S. Kulkarni, C. Rosenberg, H.-K. Chai, "Analysis of a CDN-P2P Hybrid Architecture for Cost-Effective Streaming Media Distribution", Multimedia Systems, vol. 11, no. 4, pp. 383-399, 2006

[30] M. Tran, W. Tavanapong, "Peers-assisted Dynamic Content Distribution Networks", IEEE Conference on Local Computer Networks , Sydney, Australia, 2005

[31] H. Yin, X. Liu, T. Zhan, V. Sekar, F. Qiu, C. Lin, Hui Zhang, B. Li, "LiveSky: Enhancing CDN with P2P", ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), vol. 6, Issue 3, August 2010

[32] N. Laoutaris, P. Rodriguez, L. Massoulié, "Echos: edge capacity hosting overlays of nano data centers", SIGCOMM computer communication Review 38 (1),(2008)

[33] J. He, A. Chaintreau, C. Diot, "A performance evaluation of scalable live video streaming with nano data centers". Computer Networks 53(2): 153-167, 2009

[34] M. Cha, P. Rodriguez, S. Moon, J. Crowcroft, "On Next-Generation Telco-Managed P2P TV Architectures", Proc. the 7th International Workshop on Peer-to-Peer Systems (IPTPS '08), Tampa Bay, Florida, USA, 2008

[35] V. Valancius, N. Laoutaris, L. Massoulié, C. Diot, P. Rodriguez, "Greening the internet with nano data centers", Proceedings of the 5th international conference on Emerging networking experiments and technologies (CoNEXT '09), Rome, Italy, 2009

[36] A. Nafaa, S. Murphy, L. Murphy, "Analysis of a Large-Scale VOD Architecture for Broadband Operators: A P2P-Based Solution", IEEE Communications Magazine, vol. 46, issue 12, 2008

[37] B. Krishnamurthy, J. Rexford, "Web Protocols and Practice: HTTP/1.1, Networking Protocols, Caching, and Traffic Measurement", Addison-Wesley, Reading, USA, 2001

[38] M. Rabinovich, O. Spatschek, "Web caching and replication", Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 2002

[39] S. Podlipnig, L. Böszörmenyi, "A survey of Web cache replacement strategies",ACM Computing Surveys (CSUR), Volume 35 Issue 4, December 2003

[40] M., Abrams, C. R. Standridge, G. Abdulla, S. Williams, E. Fox, 1995,"Caching proxies:

Limitationsand potentials", In Proceedings of the 4thInternational World Wide Web Conference, 1995

[41] C. C. Aggarwal, J. L.Wolf, P. S. Yu, "Caching on the World Wide Web", IEEE Trans.Knowl. Data Eng. 11, 1 (Jan.), 94–107, 1999

[42] S. Hossini-Khayat, "Investigation of generalized, caching". Ph.D. dissertation, Washington, University, St. Louis, MO, 1997

[43] M. F. Arlitt, R. J. Friedrich, T. Y. JIN, "Performance evaluation of Web proxy cache replacement policies", Tech. rep. HPL-98-97(R.1), HP Company, Palo Alto, CA, 1999.

[44] C.-Y. Chang, T. McGregor, G. Holmes, "The LRU* WWW proxy cache document replacement algorithm", In Proceedings of the Asia Pacific Web Conference, 1999

[45] J.-M. Menaud, V. Issarny, M. Banatre, "Improving effectiveness of Web caching", in recent Advances in Distributed Systems. Lecture Notes in Computer Science, vol. 1752. Springer-Verlag, Berlin, Germany, 375–401, 2000

[46] R. Salgado, D. Négru, (eds.) et al., ICT ALICANTE, Deliverable D8.2, "Pilot Integration", http://www.ict-alicante.eu/validation/download/work-package/alicante_d8.2_pre.pdf, September 2013.

[47] Y. Zhang, D. Négru, (eds.) et al., ICT ALICANTE, Deliverable D8.3, "Trials and Validation", http://www.ict-alicante.eu/validation/download/work-package/alicante_d8.3_pre.pdf, September 2013.

[48] G. Gardikis, L. Boula, G. Xilouris, A. Kourtis, E. Pallis, M. Sidibé, D. Négru "Cross-layer Monitoring in IPTV Networks", IEEE Communication Magazine, Vol. 50, Issue 7, July 2012.

[49] G. Gardikis, G. Xilouris, K. Sarsembagieva, A. Kourtis, M. Sidibe, D. Negru, Y. Chen, "A Network Cost Provision framework for Network-Aware Applications", submitted to Elsevier Computer Networks Journal, 2013.

[50] Michael Grafl, Christian Timmerer, Hermann Hellwagner, George Xilouris, Georgios Gardikis, Daniele Renzi, Stefano Battista, Eugen Borcoci, and Daniel Negru, "Scalable Media Coding enabling Content-Aware Networking", IEEE MultiMedia, IEEE Computer Society, vol. 20, no. 2, pp. 30-41, June 2013.

[51] Willy Aubry, "Conception and implementation of an hardware accelerated video adaptation platform in a home network context", PhD Thesis, University of Bordeaux 1, http://www.theses.fr/2012BOR14678, France, December 2012.

[52] ARDHMAN ANR French Project, http://ardmahn.org/php/index.

Chapter 5

Conclusions and Future Works

The overall proposed Networked Media Ecosystem towards the Future Media Internet comprises the User Environment, the Service Environment, the Network Environment, the Home-Box layer, the CAN layer, the Cross-layer Monitoring System and the Adaptation System. By providing innovative features within each[3] and by enabling efficient cooperation between them, the goal is to propose advanced features to each stakeholder of the Media Content value chain. Therefore, as a conclusion, we shortly summarize the potential impact of the proposed solutions on the industry. Then, we list our contributions to the scientific community. And finally, we outline some future works directions.

Impact on the industry

An open context framework for managing the information on users and their environment has been proposed, modeled as a User Profile instance. This system is reusing widely adopted metadata standards. It can easily support existing systems and the context models are extensible to support future systems Service Providers might want to deploy.

We developed and integrated a set of tools for monitoring key parameters at user side (i.e., QoS/QoE) that are relevant for Media Services, including device hardware, network resources, session information, etc. As a part of Cross-layer Monitoring system, these tools can be reused for future systems and deployments by Network and Service Providers, but also Equipment Vendors.

We conceived and developed a new media-centric Home-Box that features advanced functionalities. These advanced functionalities include native service provision or replication, service adaptation and redistribution. The Home-Box software modules reuse standardized protocols (e.g., for service discovery, delivery) recommandated by the related working groups (e.g., HGI, ETSI TISPAN). Thus, they can be deployed to support existing systems, can be enhanced with new functionalities (context-aware media library) and can support gradual deployment of ALICANTE-enabled systems by Service/Content Providers or Equipment Vendors.

[3] In this document, we focused on the Context-Aware Framework (Chapter 3) and the Home-Box Assisted Content Delivery Network System (Chapter 4). The other important contributions of the overall architecture are summarized in Annex A, B, and C.

We developed and integrated a set of tools for monitoring of Home-Box Distance and Home-Box resources, helping for best peer/server selection for service delivery. As a part of Cross-layer Monitoring system, these tools can be reused for future systems and deployments. Also, we integrated the results of existing work (P2P Engine based on P2P-Next[4] module, SVC technologies, etc).

We designed and implemented modules for the other parts of the architecture in the CAN layer, Service Environment or for the Adaptation System. A complete demonstrator has been set-up within the ALICANTE project and many experiments, tests and field trials have been performed. All the information on the evaluation and validation process, including all the results can be found in ALICANTE deliverables D8.1, D8.2 and D8.3 (at http://www.ict-alicante.eu/validation/delivrables/).

Contribution to the scientific community

The **Figure 31** depicts all the contributions made during the past years. The overall Network Media Ecosystem Architecture is depicted and most relevant scientific contributions are highlighted, in red for the collaborative research projects, in orange for the PhD thesis supervised and in yellow the most important research papers. Further details can be found in the Curriculum Vitae, in Annex.

[4]P2P-Next is a FP7 European project. http://www.p2p-next.org/

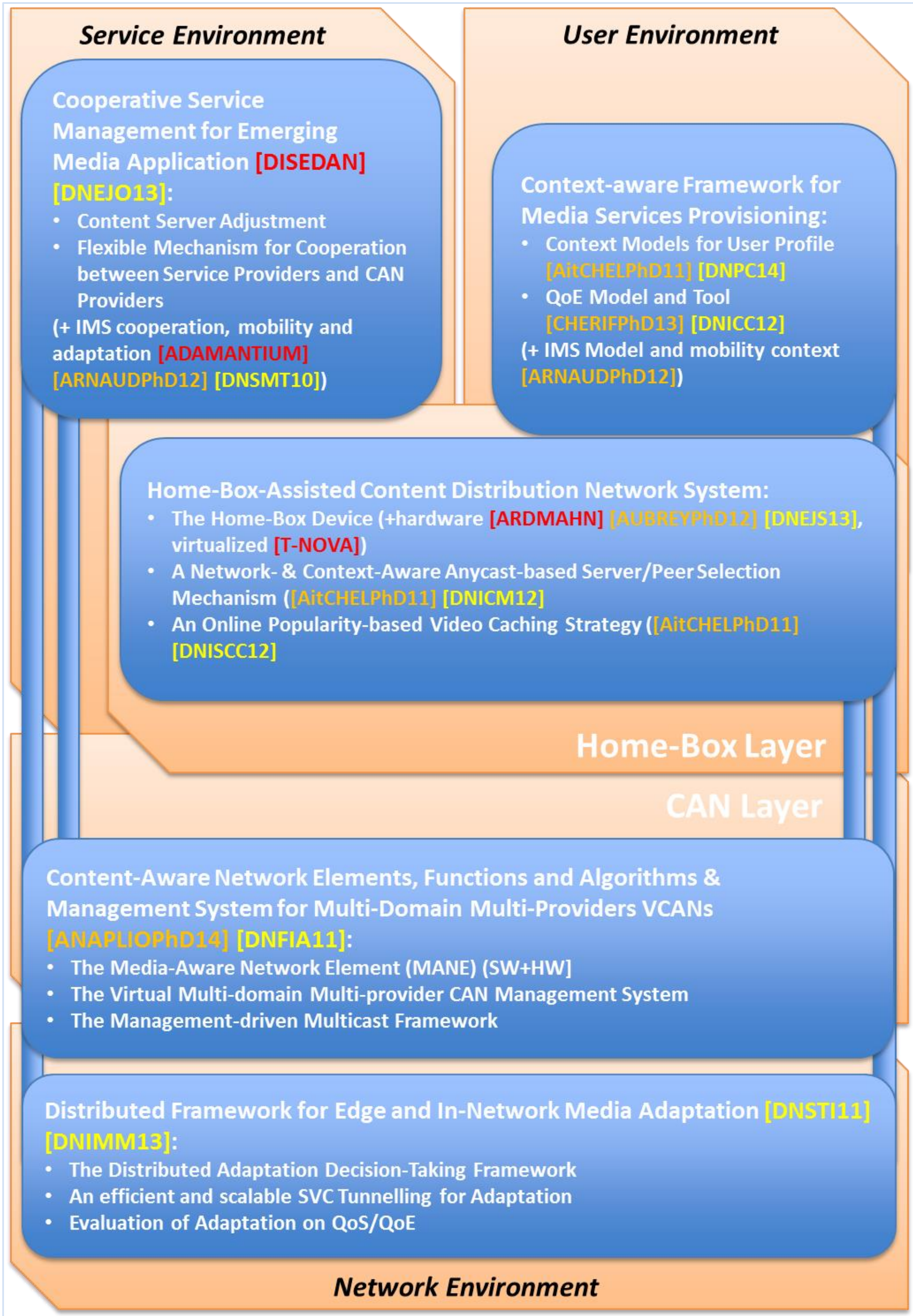


Figure 31: Summary of Research Contributions to the Scientific Community

Future Work and Perspectives

First, in the domain of Context-Awareness, device/network heterogeneity handling and user/service mobility are still evolving into challenging topics. Among future works in the Context-Awareness field, we identify the following: 1) extending the context acquisition tools, in order to complete the User Profile with other context sources as home automation; 2) exploring the context-aware adaptation scenarios to enable more interactive and intelligent Media Services, with advanced ways of content adaptation.

Concerning content distribution and overlay networks, they are in the timing of profound evolution with many challenging and fast evolving research topics. Future research directions are foreseen: 1) enhanced transport protocols for multimedia content (for example, swift-like multi-party protocols for HTTP/DASH, considering combination of HTTP-like and P2P-like protocols together); 2) social-aware overlay that takes the users' social connections into account for optimized content distribution; 3) better synergy among Data center, Content distribution networks and other distributed collaborative overlay. Along with the continuous effort towards these directions, results achieved by the project will be further exploited by telco partners for such new deployments.

Indeed, the perspectives in the domain of content adaptation considering the user and service–based information have already strongly progressed and we are in the phase of starting a new research-oriented project in the domain (starting date: 1 January 2014), within the CHIST-ERA[5] work programme, called DISEDAN (Service and User-based Distributed Selection of Content Streaming Source and Dual Adaptation). This project is part of the Context- and Content-Adaptive Communication Networks call topic. It proposes an evolutionary solution to enhance the content delivery via Internet. The project is focused on research in the area of multi-criteria content source (server) selection, considering user and server contexts and requested content and extends fundamental investigation on multi-criteria optimization applied to the selection of content streaming source. DISEDAN is in tune with the research provided in Information Centric Networks (ICN) but it is an evolutionary solution for content distribution, on the contrary of Information Centric Networks, which propose forthcoming architectures.

Another approach in the Networked Media domain is oriented to Cloud paradigm. Many approaches in media streaming today start considering evolving towards this solution. Our future work related to the Home-Box has already envisioned a solution where the Home-Box could be used and parameterized according to the developer's needs, all embedded in a virtual machine instance, hence creating a virtualized Home-Box. This way, the virtual Home-Box can

[5] Information about Chist-era work programme can be found at the following link: http://www.chistera.eu/projects-call-2012

serve as main basis for developing the future evolutions of Home Gateways, in addition to be used as is for tests and deployments of new configurations. The different aspects that will constitute the virtualized Home-Box could be: virtualization of home networking aspects; virtualization of operator's network aspects; virtualization of overall management; virtualization of Media streaming & adaptation features; virtualization of *-aware features. The virtual Home-Box will hence have the ability to serve as main basis for Home Gateway manufacturers to create their next generation Home Gateways or to validate their networking configurations and/or new features involving this equipment. This work is foreseen to be elaborated in collaboration with Viotech Communications, French SME, and within a FP7 ICT European Project called T-NOVA, starting from end 2013.

In the domain of networking and its evoluation towards media, current approaches include CAN, CON, SON, CCN, ICN (as mentioned in Chapter 1). More recently, a strong focus from all research authorities also pushes the concept of *Software Defined Networks (SDN)*. Our proposed architecture for a Networked Media Ecosystem (NetME), and materialized through the ALICANTE project, seems to be significantly in line with recent approaches like *Software Defined Networking (SDN)* and *Software Defined Internet Architectures (SDIA),* having significant support from both research and industry. The CAN architecture we propose is similar in many concepts to very recently proposed SDN architectures, potentially exposing the SDN benefits, although not following full specifications of SDN API and not using OpenFlow protocol. This creates direct and very feasible possibilities to evolve towards a full SDN solution. Among the similarities, we can mention: the Control and Data Plane are separated/decoupled, thus giving possibility to flexible changing control applications, like in SDN; network intelligence is (logically) centralized in [CAN Manager + Intra-domain Network Resource Manager] which plays the role of the SDN controller for a network domain; MANE edge routers and interior core routers are equivalent to SDN forwarders. The possibility of a multi-domain logical network governed by several cooperating "SDN controllers" is also given; in SDN, the network appears to the applications and policy engines as a single, logical switch. In our proposal**,** the network appears at higher layers as a set of parallel planes VCANs; we propose a simplified network abstraction that can be efficiently programmed: the VCANs are seen at abstract way; they can be planned and provisioned independently of the network technology.

Therefore, a work has been started conjunctly with other European research partners to investigate this possibility of evolving the proposed architecture into a SDN-related one and investigate the possibilities to go further.

Annex A: Cooperative Service Management for Emerging Media Application

Abstract **The paper presents innovation research areas in the so-called Service environment inside the EU FP7 ALICANTE project. We describe four outstanding technical concepts which are: improved service management, approach for separate phases of service and endpoint discovery, and a new system for adjustment of conten servers with seamless handover.**

***Keywords*— Service Environment, Innovation Areas, Service discovery, Service adaptation**

I. Introduction

The paper presents innovation areas in the Service environment of EU FP7 ALICANTE project [1]. In this architecture, Network, Service and User environments collaborate in order to improve the delivery of emerging Media applications' streams considering the goals of all the different actors involved. Specifically, through gaining Network- and User Context-awareness, Service Provider (SP) offers adapted service/content to the End User (EU) and facilitates load balancing to the Network Operator.

We outlined five main processes for Media content delivery in ALICANTE: (1) service publication, (2) service and content discovery, (3) content request, (4) content delivery and (5) service adaptation and adjustment of content servers. In some of them as well as in the overall Service Management we identified notable innovation research issues related to the Service Environment, which are detailed in the next sections.

II. Service Management

The targeted **technical limitations** concern the service management for content creation, publication, dissemination and context-aware adaptation.

Research challenges are flexible and integrated content and associated Service Management (SM) with *context-aware*and *network-awarecapabilities*, and also *facilities for the addition of new services*.

Key innovations are new Service Management (SM) systems with ergonomic user interface to enable Service Provider to activate content delivery services with special planning functionalities of network resources provisioning. Such a SM user interface allows the Service Provider to interact with the Content Aware Network system of ALICANTE in order to manage at the service level the management of VCANs. SM integrates also continuous monitoring of network and user environment. A complementary innovation of the SM concerns massive on-demand content processing. This processing enables x-to-svc transcoding of input source content available in multiple fragmented formats of the market, but also dynamic video editing mechanism to insert external visual objects into the original content when this is being consumed. A critical issue exists in the interoperability of serving advertisements ("ads") into multiple VoD services: The expectation of “ads” standardization activity [20] in rendering “ads” content received from multiple “ads” servers in multiple formats is too high compared to the start-of-art of available players. The implementation of the content publishing module in SM provides a solution for this issue. Based on the user profile management in ALICANTE, this is also possible to personalize the insertion of advertisement. This processing with on-demand transcoding in SM provides a publishing solution to deliver seamless content libraries of existing video on demand services into the ALICANTE architecture.

During the course of the project the SM capabilities listed above have been enhanced with the capability to perform content re-provisioning based on the underlying system conditions. This was enabled by the integration of SM software with the Content Server Adjustment (CSA) module that is able to trigger in SM additional content copies creation process when such an action is found required.

The research outcome will be a prototype of a Service Management system compliant with Service Oriented Architecture (SOA) and integrated in the Service Provider subsystem of the ALICANTE. This will be presented in [ALI7].

III. DISCOVERY: SEPARATION OF SERVICE AND END-POINT DISCOVERY

The Discovery of the service inside the ALICANTE architecture is a complex issue that takes into account not only the traditional service discovery techniques but provides a better and clearer service selection function. Specifically in traditional architecture the service discovery is an issue only of the service provider [3], which does not take into account network or user profile characteristics [4].

The **key innovation** in the ALICANTE System is that the separation between service discovery (selected by the EU) and end-point discovery (selected by the Home-Box[6] on the basis of service provider selection) allows a better selection of the delivery end-point since the system takes decisions based on passive network measurements and user profile characteristics.

More specifically when a service discovery query with related keywords (sent by the EU) arrives at the Home-Box (HB), the following modules and functions at the HB perform the service discovery process:

- Service Manager performs EU Access Control.
- The Discovery Manager is the middle point between the Home-Box and the Service Registry (SR) at the Service Provider side.
- The Arrangement Manager at the Home-Box provides means for end-points selection and generates context-aware descriptions. The Home-Box Monitoring Manager retrieves monitoring data from the HB and lower layers. Specifically, it aims at: (1) collecting data related to Home-Box resources and processed multimedia sessions (and updating the User Profile accordingly) and (2) requesting and processing Network Distance data as provided by the Content-Aware Network (CAN) layer. The query arriving from the Arrangement Manager to the HB Monitoring Manager contains the candidate end-point and also the service to be provided. These two parameters together with the IP addresses of the querying Home-Box are sent to the CAN Manager in order to retrieve the Network Distance data, whenever the corresponding data are not in the cache. The Network Distance (published in [ALI6]) is used for both server selection and peer selection in ALICANTE as it is amply described in Innovation Area 2.

This is also the approach of other CAN architectures. In fact, we can say that this feature characterizes the Content Aware Networks: the possibility of including real-time measurements to describe the different entities in network and service layers in order to improve the delivery of the content. Together with COMET partners we introduced a novel approach for introducing information of network and servers [11] in content source selection. Unlike previous works that are focusing on one parameter such as latency [12] or path length [13], the proposed decision algorithm follows the multi-criteria optimization approach, where long-term scale (dimensioning) and short-term scale (measurements) metrics are taken into account.

The cooperation with COMET included the investigation of new approaches to introduce server awareness in the end-point selection process. Till now, the main investigated approaches assumed that decisions could be taken by the network infrastructure, by the client applications, or by the content provider. Among solutions relaying on the network infrastructure, we can distinguish the "route-by-name" [14][15] and DNS-like approaches. The "route-by-name" approach [15] assumes that every content-aware node forwards the content request towards the destination server based on its local knowledge. In these Data-oriented network architectures, the decision about server selection is taken in distributed way as a concatenation of local optimizations. Therefore, the final solution may not be optimized in the global scope. On the other side, the "DNS-like" approaches [16] collect information about available content replicas, content server status and network conditions, then use it for selecting the best content server to serve consumer request. In principle, the DNS-like approaches are centralized and could lead to the globally optimal solution. However, the challenge for the DNS-like approaches is to design effective and scalable information system which collects information about content localisation, server load and network status with appropriate accuracy. The investigated approaches exploit distributed information systems designed on federation principles.

The client side decision strategy assumes that the application selects the best content based on information collected by itself. The investigated approaches [17][18] exploit the dynamic probing and statistical estimation of different information such as round trip delay, bandwidth, servers responsiveness. The results presented in [17] confirm that even simple dynamic probing outperforms blind client-side

[6] Home-Box is present in each EU's home and is the 'meeting point' for the three environments (user, service provider and network provider), enabling collaboration.

approaches. However, the main limitation of client side strategy is its limited scalability in an Internet-wide deployment.

In order to overcome these limitations, the server side selection strategy had been investigated [18][19]. It allows to aggregate information at the server side and to reuse it for redirecting the content requests coming from different specific areas. Scalability requirements for server selection strategies indicated the need for some offline pre-computation as a preparation to serve content requests in real time.

This is the approach investigated together with COMET researchers and it based on the two-phase approach for content source selection. [ALI4] presented an improved content source selection algorithm with weighted factors of server and network awareness, whereas [ALI5] introduced an evolutionary multi-objective optimization algorithm in CANs for two-phase selection process.

- At last, the Delivery Service Manager is responsible for media service delivery from the Home-Box to the End-User.

In summary, the two level filtering approach of the ALICANTE System is providing: (a) service filtering in the Service Provider based on the service query and (b) service filtering in the HB (Arrangement Manager) on the basis of more information such as User profile, terminal and HB capabilities and Network Distances. The **target research outcomes** are the following:

1. Network features embedded in service discovery.
2. User data characteristics (i.e. CPU, Display, etc.) embedded in service discovery.
3. Enhanced traditional service discovery techniques.

IV. ADJUSTMENT OF CONTENT SERVERS

Adjustment of content servers enables to lower the probability of congestion and overload situations in the network as well as dealing with such situations if they occur. It brings benefits to all the parties involved in Media delivery. For the End Users it results in a higher Quality of Experience of the received media content, the Service Provider is able to ensure better service increasing clients' satisfaction, and the Network Provider achieves reduction of overloads within the network. In order to provide efficient content server management, we specified, implemented and validated new approaches in the **research areas** of fault location, adjustment of content provisioning as well as handover of media streams.

At first, we proposed an integrated system for adapting the content servers in [ALI1] which components were later on mapped to ALICANTE System architecture. The system is able to adjust content provisioning and improve the content server selection process; however, the **key innovation** of the system is the **seamless handover mechanism** of on-going connections that allows for on-the-fly switching of content source without noticeable disruptions in the content delivery.

Several software components were specified and implemented from scratch including the Content Servers Adjustment (CSA) module at the Service Management for network fault localisation and centralized adjustment decisions making and Seamless Handover and DASH Proxy modules at the Home-Box to enable stream handover.

Whenever the CSA receives real-time notifications indicating that one or more connections are transmitting with poor quality, it identifies and localises the causes (e.g., network congestion or content server overload) at the global view of all affected content sources and content sinks, and provide suggestions for possible solutions. The action triggered by the CSA may include, e.g., (1) providing a list of content servers and/or paths that should be avoided during content delivery, (2) request for creating and placing in the network additional content copies for load balancing purposes or (3) request for given existing connections handover. Adamant [8] is a system for Sub/Pub networks which actually provides similar functionalities, but the actions provided deal with on the transport protocol.

The CSA fault localization algorithm (FLA) specification was presented in [ALI1] and the results of conducted FLA accuracy and performance simulations were presented in [9]. The considerations related to CSA decision process were provided in [9][10] and complemented with simulation studies on CSA operational parameters optimization described in [10].

Although seamless handover is conceptually simple, the technical details cover content stream encapsulation (e.g., stream containers), network transport protocols (e.g., TCP and UDP) and application session protocols (e.g., RTSP, HTTP, IGMP). The **State of the Art** of seamless handover is very extensive (e.g.,[2][5][6][7]), but these solutions propose the handover at the client side and have other purposes such as user mobility issues. The proposed handover is performed at the server side which allows for an on-demand switching of the source content server for improving the transmission.

The technical aspects and implementation details of the seamless handover mechanism for HTTP and RTSP/RTP protocols were presented in [ALI1] We have also conducted series of test measurements in local testing environment, presented in [ALI1] in order to demonstrate the performance of the handover solution itself and the benefits of its introduction into a content delivery system.

In [ALI2] and [ALI8], we presented the introduction of Seamless handover during streaming of media segments to enable dual adaptation capabilities in Dynamic Adaptive Streaming over HTTP (DASH) connections. In comparison with current Media adaptation solutions, the Seamless handover reaches higher Quality of Experience in the case when the adaptation counteracts an overload situation at the server side or in the core network. In order to take advantages of both Media and Seamless handover functionalities, we proposed to carry out the last one (if needed) during the transmission of individual segments, i.e., in-segment Seamless handover, whereas Media adaptation will be performed (if needed) at the boundary of each segment, as specified in the DASH standard. We implemented and integrated the proposed solution in DASH library and performed tests in order to demonstrate the advantages of dual adaptation solution. The final solution (mathematical analysis as well as test results) adopted in ALICANTE for integrating Media adaptation and Seamless Handover into one unified solution will be presented in [ALI3].

The Seamless Handover and DASH Proxy modules were fully integrated with the Home-Box environment components and take benefit from the implemented mechanism for e.g. stream monitoring and new server selection.

V. CONCLUSIONS

The innovation areas described in this paper results in challenges for improving the publication and delivery of content for emerging media applications. This area will become a key element during the development in the future Internet.

REFERENCES

[1] EU FP7 ALICANTE project homepage: http://www.ict-alicante.eu/

[2] E. Piri, K. Pentikousis “IEEE 802.21”,The Internet Protocol Journal, Vol.12, No.2, June 2009.

[3] V. Deora, J. Shao, G. Shercliff, P.J. Stockreisser, W.A. Gray, and N.J. Fiddian,”Incorporating QoS Specifications in Service Discovery” Web Information Systems – WISE 2004 Workshops (2004), pp. 252-263

[4] C. Ververidis, G. Polyzos, ”Service Discovery for Mobile Ad Hoc Networks: A Survey of Issues and Techniques”. Communications Surveys & Tutorials, IEEE, Volume: 10 Issue:3 2008

[5] “Retail Wireless Networks Validated Reference Design. Version 3.3”. Aruba networks Solution Guide. December 2008.

[6] ITU-T Recommendation H.323: “Packet-based multimedia communications systems”. (11/2000)

[7] NetMotion Wireless Company, “Comparing Mobile VPN Technologies. White paper”. © 2009.

[8] J. Hoffert, A. Gokhale and D. Schmidt, “Autonomic Adaptation of Publish/Subscribe Middleware in Dynamic Environments”. Nashville, Vanderbilt University, 2011.

[9] J. Mongay Batalla (ed.) et al., ICT ALICANTE, Deliverable D5.4.1, "Service and Provisioning Adaptation Module – Intermediate", August 2012.

[10] F. Bouilhaguet (ed.) et al., ICT ALICANTE, Deliverable D5.5, “The ALICANTE Service Environment – Final”, July 2013.

[11] A. Beben, J. Mongay Batalla, W. Chai, and J. Sliwinski, “”Multi-criteria decision algorithms for efficient content delivery in content networks”. Springer annals of telecommunications - annales des télécommunications: Volume 68, Issue 3 (2013), Page 153-165. http://www.springerlink.com/openurl.asp?genre=article&id=doi:10.1007/s12243-012-0321-z

[12] S. Ratnasamy, M. Handley, R. M. Karp and S. Shenker, ‘Topologically-Aware Overlay Construction and Server Selection”. IEEE INFOCOM, 2002

[13] N. Magharei and R. Rejaie, “Prime: Peer-to-peer receiver-driven mesh-based streaming”, IEEE/ACM Transactions on Networking, 2009

[14] T. Koponen, M. Chawla, B-G. Chun, A. Ermolinskiy, K. H. Kim, S. Shenker and I. Stoica, “A Data-oriented (and Beyond) Network Architecture,” in Proc. ACM SIGCOMM ’07, Kyoto, Japan, Aug. 2007.

[15] “NDN – Named Data Networking project,” http://www.named-data.org/

[16] EU 7th Framework Program Scalable and Adaptive Internet Solutions (SAIL) project. http://www.sail-project.eu/.

[17] S. G. Dykes, K. A. Robbins and C. L. Jeffery, “An Empirical Evaluation of Client-side Server Selection Algorithms,” Proc. IEEE INFOCOM ‘00, Tel-Aviv, Israel, March 2000.

[18] E. Zegura, M. H. Ammar, Z. Fei and S. Bhattacharjee, “Application-layer Anycasting: A Server Selection Architecture and Use in a Replicated Web Service,” IEEE/ACM Trans. on Netw., vol. 8, no. 4, Aug. 2000, pp. 455-466.

[19] V. Lenders, M. May and B. Plattner, “Density-based Anycast: A Robust Routing Strategy for Wireless Ad Hoc Networks,” IEEE/ACM Trans. on Netw., vol. 16, no. 4, Aug. 2008, pp. 852-863.

[20] VAST (vast video ad serving template) specification - www.iab.net/media/file/VASTv3.0.pdf

RELATED ALICANTE RESEARCH PAPERS AND PUBLICATIONS

[ALI1] J. Mongay Batalla, S. Janikowski, P. Krawiec, B. Belter, L. Lopatowski and J. Gutkowski, Dynamic Seamless Handover of Video Servers”. Proceedings of The 12th International Conference on Telecommunications ConTEL 2013. Published by IEEE Computer Society. Zagreb (Croatia). June 26-28, 2013.

[ALI2] J. Mongay Batalla and S. Janikowski, “In-Segment Content Server Adaptation for Dual Adaptation Mechanism in DASH”. Proceedings of Fifth International Conference on Computational Intelligence, Communication Systems and Networks CICSyN 2013. Published by IEEE Computer Society. Madrid (Spain). June 5-7, 2013.

[ALI3] IN PREPARATION J. Mongay Batalla, P. Krawiec et al., “Content source adaptation in conjunction to Media adaptation for improved content delivery”. **In preparation**. It will contain the analysis of seamless handover provided in D5F and test results.

[ALI4] J. Mongay Batalla, A. Beben and Y. Chen, “Optimization of decision process in Network and Server-aware algorithms”. ISBN: 978-1-4673-1391-9/12 Proceedings of 15th International Telecommunications Network Strategy and Planning Symposium, Networks 2012. Published by IEEE Computer Society. Rome (Italy). DOI: 10.1109/NETWKS.2012.6381685. October 2012.

[ALI5] J. Mongay Batalla, A. Bęben and P. Krawiec, "EMO algorithm for two-phase selection process in Content Aware Networks". **Submitted** to Special Issue of the European Journal of Operational Research on "Evolutionary Multiobjective Optimization". September 2013.

[ALI6] G. Gardikis, L. Boula, G. Xilouris, A. Kourtis, E. Pallis, M. Sidibe, D. Negru, Cross-layer Monitoring in IPTV Networks, IEEE Communications Magazine, Vol.50(7), June 2012, pp.76-84

[ALI7] IN PREPARATION J. Mongay Batalla, F. Bouilhaguet, Ł. Łopatowski, S.Obreja, Z. Leib, "Advanced Service Management for Content Aware Networks". **In preparation**.

[ALI8] J. Mongay Batalla, P.Krawiec, A.Beben, "Mechanizm Wewnątrzsegmentowej Adaptacji Żródła Strumienia MPEG DASH", Gdańsk, KSTiT September 2013. To appear in Przegląd Telekomunikacyjny, Sigma eds..

Annex B: Content-Awareness and Management System for VCANs

***Abstract* — This document presents a summary of the innovations of the EU FP7 ALICANTE, WP6 "Content-Aware Network (CAN) Environment". Two specific areas are discussed: *content awareness* realization through functions and algorithms executed by new network elements; a novel*management architecture* and the associated system, to govern multi-provider, multi-domain virtual CAN construction, offering and utilization by Service Providers and End Users. Problem statement and main research challenges are shortly discussed and then the ALICANTE beyond state of the art (i.e. innovations) and scientific/technical results are presented.**

***Keywords* — Content-Aware Networking, DPI, traffic classification, machine learning, Network and services management, virtualization, Service Level Agreements, Quality of Services.**

I. Introduction & Problem Statement

It is largely accepted that Future Internet (FI) will be strongly service and content- orientated. The need to support a large multimedia e-services, including new end-users role in consumption/generation of content, while having access to various QoS enabled services, asked that network architecture exploit content-awareness. On the other side, it is recognized that current Internet has serious limitations expressed in so called "ossification".

Trying to respond to media/content orientation and also to overcome some of the current Internet limitations, the EU FP7 **ALICANTE** project proposes an evolutionary approach based on *Content Aware Networks (CAN)* and Network *Aware Applications (NAA)* concepts. Two major directions have been investigated in WP6:

- how to achieve ***content awareness (CA) at network level*** in order to optimize media flows transport and delivery;
- how to create a ***CAN management architecture,*** flexible, evolvable,capable to lightly virtualize and logically split the network resources and offer them to as abstract logical slices to the upper layers.

In order *to realize CA*, the first problem to be solved is traffic identification. Currently there are many traffic identification techniques, categorized according to the method used to extract information from the incoming data flows (e.g. analyzing headers, L3 - L7). Other methods combine multilayer information extracted through application data inspection (i.e. Deep Packet Inspection -DPI), and others - less invasive but more complex- are based on statistical flow information. The ALICANTE CA solution creates a framework to allow legacy and future network providers to dynamically exploit any of the available traffic identification techniques, and apply them specifically on media flows. When QoS assurance is required, CA(N) actually represents the ability of network elements to identify the incoming data flows, classify them according to specific flow/content characteristics and treat them in a specific manner (individually or in groups) by dynamically applying specific policies according to each flow's QoS requirements.

At *overall system level,* the CAN (including IP network layer) Environment aims to enhance cooperation between network and higher layer as to offer to the services and applications layers a CA customized transport, at network layer. Towards this, one should create a flexible and dynamic *CA-oriented management and control (M&C)* framework, in which multiple actors (Service Providers, Network Providers, Content Providers, End-Users) can cooperate for multimedia services delivery over large IP multi-domain network infrastructures. The media transport is performed by a *novel virtual CAN overlay*, intelligently supporting multimedia content distribution. The CAN layer offers (based on novel Media Aware Network Elements- MANE) to the higher layers CA customizable connectivity services and cooperates with upper service layer – which, in its turn, can be network-aware.

Section 2 presents the main research challenges, w.r.t WP6 research activities. Note that state of the art, and related work have been discussed in details in Deliverables D2.1, D6.1.1, D6.2.1, D6.x and aggregated in D6F (chapter 2).

Section 3 is the main one, presenting the major WP6 innovations beyond SotA .

First, innovations in the CA area are summarized. The ALICANTE approach has been to combine state-of-the-art traffic classification techniques and their integration inside a new Media Aware Network Element (MANE) capable to recognize and also process the media flows in specific way, conforming their types and QoS requirements.

Second, an innovative CAN architecture is presented, evolutionary and evolvable, defined for CAN and network layers, to support the construction of virtual customizable CANs, oriented to multimedia flows transportation. A management system has been defined, designed and implemented having as objective to federate the construction and control the operation of multi-domain, multi-provider VCANs, where forwarders elements are MANE.

Section 4 summarizes the scientific and technical results.

Details on the topics presented here are collected together in Deliverable D6F. Several papers (see D6F Reference list) have treated and developed specific topics.

II. Research Challenges Summary

A detailed discussion on state of the art, existing solutions, their limitations and research challenges is included in Deliverable D2.1, D6x – intermediate and D6F. Here a summary only is included.

The research challenges w.r.t Content Awareness includes:

- Creating a classification solution capable of achieving acceptable levels of accuracy and completeness for a variety of different applications (focusing on media services), networks and time periods.
- Studying the effect on the traffic profile and dynamics of a certain flow during transmission and how this affects the identification accuracy.
- Selection of the appropriate interval at which the CATI information (in the case of explicit signaling using CATI) will be inserted in order not to introduce unnecessary overhead.
- The assessment of the performance of the traffic identification algorithms against various types of services, flow bandwidth and durations etc.
- Traffic profiling of ALICANTE multimedia service flows in order to be used as signature in matching algorithms.
- Integration of state-of-the art traffic classification algorithms and ALICANTE specific ones (i.e CATI, DPI, etc.) inside a Media Aware Network Element capable of exploiting the identification capabilities in order to provide QoS to the respective media flows.
- Develop a new type of forwarder (MANE) integrating the CA and process the media flow accordingly (including with QoS guarantees), under coordination of external M&C Plane, finally capable to slice the network transport resources in the Data Plane in customizable way at request of Service Provider. MANE should work, at request, in both unicast or multicast mode.
- Develop at network level (MANE and core routers) of a flexible solution capable to treat in selective way the media flows under different levels of connectivity services (Fully Managed -FM, Partially Managed-PM or Unmanaged (best effort- UM).
- Study the scalability of MANE, given the additional processing tasks w.r.t. traditional router; identifying the best tradeoffs - complexity performance and bottlenecks for various design solutions (HW-based, SW based).

Research challenges related to Management and Control, are summarized below. Some of them are general (i.e. not linked directly with CA) and others are related to the Content Awareness general framework.

General research challenges:

- To define and develop a flexible CAN Management architecture (at CAN layer) capable to offer on demand, to multiple Service Providers, enhanced connectivity services, (in unicast and/or multicast modes) organized in virtual (logical) slices, separated between different SPs.
- The CAN Management architecture should be scalable in terms of number of SPs or network domains involved.
- Define an architecture which allows to obtain logical aggregated/abstracted view on the logical network slices, while preserving freedom in resource management for infrastructure providers (Network Providers).
- SP should be not burdened with task of constructing the logical transport infrastructure, but they only ask and use it. The network infrastructure should be hidden to SP.

The relationships of SP with CAN layer should be dynamic SLAs (based on negotiation protocol).

- Define an integrated management architecture which allows the transport services to span single or multiple domains. Integration here means coordination of activities like planning, negotiation, provisioning, offering, operation (installation, modification, monitoring, termination), while cooperating with IP network layer. Additionally, advertising, discovery functions can be added.
- Data Plane should be programmable by external control logic.

The specific research challenges of the CAN Management and control, related to the ALICANTE Content Awareness basic characteristic and cooperation inside general Alicante architecture are the following:

- Define a management framework inside a novel CAN architectural layer, capable to create on demand, install, maintain, modify, and terminate Virtual Content Aware Networks (VCANs) over single or multiple network domains.
- Customize the VCANs as to be capable to offer customized services (unicast/multicast, FM/PM or UM).
- Develop general algorithms to optimize the inter-domain routing and make VCAN resource reservation and mapping onto networking resources while respecting CA and QoS requirements.
- Development of a hybrid multicast subsystem, management driven, capable to combine in CA style, the IP multicast overlay multicast and P2P.
- Develop a powerful and hierarchical CA-oriented monitoring system capable to supervise network status and SLAs fulfillment.
- Allow cooperation between VCAN provisioning and media flow adaptation inside the VCANs both in unicast or multicast mode. Scalable Video Coding (SVC) technologies should cooperate with multicast.

III. ALICANTE CAN AREA – INNOVATIONS

A. Innovations in Content Awareness

The key innovations of the **ALICANTE** in the frame of RC3 (*Content Awareness at Network Level*) are considered to be:

- **Multi-tier configurable media awareness** via integration and extension of state-of-the-art mechanisms, giving the opportunity to recognize and classify: fully managed and partially managed **ALICANTE** services, CATI-marked media services, based on DPI and CATI parsing; unmanaged media services, based on inspection of application layer information (DPI). **Ref.**: D2.1 D6.1.1; **ALICANTE** papers:[ALI-1][ALI-4][ALI-7];
- Prototype development of a **novel network element: MANE** – that integrates legacy router functionalities with Content-Awareness capabilities **Ref.**: D2.1, D6.1.1; **ALICANTE** papers:[ALI-12][ALI-15];
- **Media Aware Traffic differentiation and forwarding**, i.e., applying the appropriate policies to each media service flow (as contracted between SP and CAN/Network Provider) in terms of: traffic differentiation and QoS assurance; in-network security; multicast/unicast routing.**Ref.**: D6.1; **ALICANTE** papers:[ALI-8][ALI-9][ALI-10];
- Flexible methods to achieve **content-aware flows classification and processing** (filtering, QoS mapping, security routing/forwarding, etc.), based on cooperation between Data Plane methods and Control Plane mechanism, on explicit signaling (CATI) and inherent application detection. **References**: D6.1.1; **ALICANTE** papers:[ALI-3][ALI-14][ALI-17].

B. Innovation in VCAN Management

ALICANTE proposes enhancements beyond SotA, related to some of the virtualization challenges presented in the previous sections as follows:

- **Define a new entity business entity - CAN Provider;** a dual-layer (CAN and IP) network management framework is defined in **ALICANTE** allowing NPs to enrich their business by cooperating with CANPs, while still preserving their independency;
- Creates **on demand logically isolated *VCANs,* realized as parallel Internet planes based on light virtualization** (in the Data Plane only), thus providing *enhanced connectivity services*(based on SLA SP-CANP) to serve various needs of Service Providers and End-Users;
- **Optimized inter-domain and intra-domain mapping of VCANs,** onto several domain network resources by using a combined algorithm for QoS constrained routing, mapping and resource reservations, based on a virtual topology service.

ALICANTE innovations related to some of the ***Management and Control challenges*** and limitations are:

- **Architecture supporting both vertical and horizontal integration SLAs** for several level of guarantees Fully/Partially/Unmanaged VCAN services. Dynamicity of I/F between SP and CANP are based on negotiation protocols (online SLAs modifications). Different security levels are supported;
- **The distributed M&C** (each domain has its Intra-NRM and an associated CAN Manager) assures large scale provisioning capabilities and integration of independent NPs, supporting CA multi-domain peering solution;
- **Flexible VCAN operations, in various modes: unicast, multicast, broadcast, P2P** and combinations, synchronous/ asynchronous, with different levels of QoS/QoE, availability, etc. Several methods are defined to achieve content-aware flows classification and processing (filtering, QoS, security routing/forwarding, etc.), based on cooperation between M&C Plane and Data Plane methods to recognise content-types.

Other M&C new features are: QoS/QoE optimisation based on: **1) CAN/NAA interaction and 2) cooperation between resource provisioning (SLA) and media flow adaptation;**hierarchical monitoring at CAN layer cooperating with the upper layers.

The M&C system governs a distributed intelligence at network layer by developing CA in new core network elements, while preserving the compatibility of inter-working with traditional network elements; a novel network node is defined: MANE which is a CA-router complemented with MPLS-DiffServ technologies. The **ALICANTE** M&C creates new flexible *business possibilities* for entities to organize and exploit their resources in cooperation with other actors: enriches NP roles by cooperation with CANPs or to become both NP + CANP; extends the NP connectivity services to multi-domain regions, based on CAN concepts; preserves the independency of business actors*;* allows incremental of VCAN deployment; supports an open architecture (VCAN services can be offered not only to **ALICANTE**-compliant SP, but also to IMS or CDN-oriented SP.

C. Innovation in Multicast

The main specific innovations brought by the **ALICANTE** multicast subsystem in the area of multicast communications are the following:

- **Technical aspects***:*
  - It combines, in a *flexible mode,* three multicast technologies: *overlay multicast* (intra or inter-domain), *IP multicast* (intra-domain where available), and *peer-to-peer (P2P)* multicast at the edges of the hybrid tree;
  - The multicast architecture is developed in CAN context. The multicast trees are multicast-enabled customized VCANs, spanning several IP network domains
  - The *multicast trees are QoS enabled* - on request of Service Providers; the multicast solution can *cooperate with Scalable Video Coding* – based *adaptation methods* in order to preserve or enhance the final *Quality of Experience* perceived by the users.
- **Business aspects:**
  - *Management-driven multicast,* thus offering to the Management Plane an overall view of the multicast infrastructure. *A new Management Driven Multicast Protocol* has been proposed in the project;
  - *Hybrid multicast solution,* multi-domain, multi-provider, thus solving the heterogeneity of IP multicast capabilities (available or not in different IP network domains);
  - *Several SPs can independently ask* to the CAN Provider(s) and then simply use several multicast trees (logically separated) by sharing the same network infrastructure;
  - *NPs are free to allocate connectivity resources*, conforming their own policies;
  - Improve the management capabilities due to *Software Defined Network similar architecture* (see below some technical aspects).

D. Flexibility and scalability aspects

The **ALICANTE** overall system flexibility and scalability essentially depend on its M&C architecture. For *VCAN planning, provisioning and exploitation:*it was adopted per-domain partially centralized solution; this avoids full centralized VCAN management (non-scalable), but allowing a coherent per-domain management. However, the initiator

CAN Manager has the overall consistent image of a multi-domain VCAN.

No per-flow signaling between CAN Managers exist. The VCAN SP–CANP negotiation is done per VCAN, described in terms of aggregated traffic trunks. The SP negotiates its VCAN(s) with a single CAN Manager irrespective if it wants a single or a multi-domain spanned VCAN.

A hierarchical overlay solution is applied for inter-domain peering and routing where each CAN Manager knows its inter-domain connections. The CAN Manager initiating a multi-domain VCAN is the coordinator of this hierarchy, without having to know details on each domain VCAN resources. The monitoring at CAN layer and network layer is performed at an aggregated level.

For *Multicast M&C,* a hybrid solution is used, with IP level multicast intra-domain wherever is possible. Aggregated trees can be constructed, usable by multicast sessions having similar QoS characteristics. The IP multicast can cooperate with P2P (used by the HBs), thus assuring a better scalability.

For *Routing and Forwarding:* the highest bandwidth paths are selected thus optimizing the network resource. The length of the paths can be minimized by using higher layer tiers domains when necessary. The length of the paths between HBs working in P2P mode, or between HBs and CSs, will be minimized by delivering network distances information to HBs to help the peering process. However, the solution is flexible in the sense that other more complex additive metrics can be used as well.

Management of configurable (on demand) types of VCANs: the amount of processing in the Data Plane affects the scalability of the system. To be flexible with respect of different SP needs, and not proceed to a very rich granularity if no need for it, the CAN layer may offer several types of VCANs seen as parallel planes. The M&C can configure the VCANs (at request of the SP), so that to offer gradual scalability properties and QoS differentiation capabilities, with different types of classes of services (see D5.1.1).

E. ALICANTE – inline with recent proposals for Internet architectures

The ALICANTE CAN architectural approach is a *middle way solution* between the content-agnostic traditional Internet and the revolutionary (“clean slate”) *Content/Information Oriented Networking (CON/ICN)*; however, it avoids serious scalability and complexity problems of the CON/ICN. It is recognized that clean slate solutions, despite their advanced architectures, claiming to solve may shortcomings of the current Internet, have not achieved a large application in the real network deployments. On the other side, the **ALICANTE** architecture is significantly inline with recent approaches like *Software Defined Networking (SDN)* and *Software Defined Internet Architectures (SDIA),* having significant support from bothresearch and industry.

ALICANTE CAN architecture is similar in many concepts to very recently proposed SDN architectures, potentially exposing the SDN benefits, although not following full specifications of SDN API and not using OpenFlow protocol. This creates real possibilities to evolve towards a full SDN solution. The similarities are listed below.

- ALICANTE is an evolutionary architecture; it can be seamlessly deployed (See D2.1. D6.2.1. , D6F)
- The Control and Data Plane are separated/decoupled, thus giving possibility to flexible changing control applications, like in SDN. The main control application comprises QoS constrained routing, resource allocation, admission control and VCAN mapping which are included in the CAN Manager. The network light virtualization is performed jointly by Intra-NRM, and CAN manager which hides the characteristics of actual network topology to the control applications and higher layers.
- Network intelligence is (logically) centralized in [CAN Manager + Intra-domain Network Resource Manager] which plays the role of the SDN controller for a network domain.
- MANE edge routers and interior core routers are equivalent to SDN forwarders. ALICANTE offers the possibility of a multi-domain logical network governed by several cooperating “SDN controllers”.
- The degree of centralization is configurable in ALICANTE by defining the placement how many controllers exist and what set of forwarders they control.

- Execute M&C SW on general purpose HW (external commodity servers – w.r.t network elements)
- Decoupled from specific networking HW **(**MANE and core routers are viewed at CAN layer in abstract way)
- Data Plane is programmable; all configurations for MANE and Core routers are determined in CAN and Network M&C and downloaded in the routers.
- The architecture defines the control for a network (and not for a network device). At the level of CAN Manager, one has an overall image on the static and dynamic characteristics of all VCANs; at the level of Intra-NRM, one has full control on the network domain associated with that Intra-NRM.
- In SDN the network appears to the applications and policy engines as a single, logical switch. In ALICANTE**,** the network appears at higher layers as a set of parallel planes VCANs
- Simplified network abstraction that can be efficiently programmed: the VCANs are seen at abstract way; they can be planned and provisioned independently of the network technology. E.g. for inter-domain routing one use a different protocol QoS enabled. Two variants of routing have been deployed.

Alicante is also open to the new trends of Software Defined Internet Architecture, which aims to go further beyond SDN. The Alicante can support SDIA-proposed large network split in core and edges. In each core domain an arbitrary network technology can be used (MPLS, IP, L2-only, etc.) while at the edges software routing is used (programmable under SDN-Like control). MANE - LINUX and ALICANTE M&C framework allow naturally such a development. More than that, ALICANTE CAN Management architecture allow development of an Inter-domain Service Model (ISM) like in SDIA, where BGP-like protocol running between controllers is used or whatever other inter-domain routing protocol.

IV. Summary of Scientific and Technical Results

The main scientific results are summarized below:

- Novel content-aware Media Aware Network Element MANE, performing media flows processing (various combinations of classification techniques, based on new CA-transport Information (CATI) included in packets and/or Deep Packet Inspection, encapsulation, QoS, adaptation, etc.) and forwarding over several network domains;
- Novel CAN layer middleware assuring abstraction of networking resources for higher layers;
- M&C framework of the CAN and Network environment - with partial centralization - (supporting multi-provider, multi-domain) - to create on demand VCANs (dynamic SLAs) and exploit them;
- SDN and SDIA similar architecture - i.e M&C controls the forwarders; programmability of Data Plane;
- Combined algorithms for resource allocation and optimization at CAN and network layer- included in CAN layer as an control application (in SDN sense)
- Network level resource management (MPLS, DiffServ) and cooperation with CAN layer;
- Flexible and multi-domain management driven hybrid multicast solution (IP-native, Overlay, P2P);
- Multi-layer monitoring plane supporting the VCAN exploitation (SLA fulfillment verification and adaptation);
- Proactive resource provisioning and media flow adaptation cooperation.

The WP6 technical achievements are summarized below:

Content/Media Awareness:

Content aware classifiers. Two MANE versions prototypes (HW and SW) developed. Identification of bottlenecks in terms of speed processing and find out a good split balance between HW/SW tasks.Content aware hierarchical monitoring system.

Virtual Multi-domain Multi-provider Content Aware Networks Management:

Flexible business model multi-provider, multi-domain. VCAN Management software for CAN Managers. Intra-domain Network Resource Manager software, cooperating with CAN Manager. Software for combined algorithm for VCAN mapping onto network resources (optimized QoS routing, admission control and resource reservation, VCAN mapping). Hierarchical monitoring plane software supporting VCANs, allowing SLA fulfillment checking and cooperation between resource provisioning and media flow adaptation. Scalability analysis.

Management Driven Multicast:

Multicast VCAN-oriented management framework software. Multicast, multi-domain, QoS enabled routing protocol software. Multicast bridge software inside MANE. Performance analysis of multicast.

V. ALICANTE Papers References

[ALI-1] H. Koumaras, D. Negru, E. Borcoci, V. Koumaras, C. Troulos, Y. Lapid, E. Pallis, M. Sidibé, A. Pinto, G. Gardikis, G. Xilouris, C. Timmerer, "Media Ecosystems: A Novel Approach for Content-Awareness in Future Networks", Future Internet: Achievements and Promising Technology, Springer Verlag, pp369-380, May 2011, (also - D2.1 related).

[ALI-2] E. Borcoci, R. Miruta, S. Obreja "Inter-domain Peering and Overlay Topologies Support for Content-Aware Networks Dedicated to Multimedia Applications", International Journal on Advances în Telecommunications v 5 n 3&4 2012;

[ALI-3] R. Miruta, C. Stanuica, E. Borcoci "New Fields în Classying Algorithms for Content Awareness",International Journal of Informaţion Technology and WebEngineering, (IJITWE)",2012 7(2), 1-15, April-June 2012.

[ALI-4] E.Borcoci, S.Obreja, et. al., "Resource Management in Multi-Domain Content-Aware Networks for Multimedia Applications" , International Journal on Advances in Networks and Services", vol 5 no 1 & 2, year 2012, http://www.iariajournals.org/networks_and_services/.

[ALI-5] R.Iorga,(UPB) António Pinto,(INESC) E.Borcoci,(UPB) G. Carneiro,(INESC) R.Miruta,(UPB) T.Calcada,(INESC) "Multicast Services and Resources Management in Virtual Content Aware Networks", paper submitted to IEEE Comm.Magazine, December 2012

[ALI-6] G. Gardikis, G. Xilouris, K. Sarsembagieva, A. Kourtis, M Sidibe, D. Negru, Y. Chen, "A Network Cost Provision framework for Network-Aware Applications", submitted to Computer Networks.

F. ALICANTE Conferences Papers

[ALI-7] E.Borcoci (UPB), D.Negru (LABRI), C.Timmerer (UNIKLU), **"A Novel Architecture for Multimedia Distribution based on Content-Aware Networking"**, Proceedings of the 3rd International Conference on Communication Theory, Reliability, and Quality of Service (CTRQ 2010), Athens/Glyfada, Greece, June 13-19, 2010, (also - D2.1 related).

[ALI-8] E.Borcoci, R.Iorga (UPB) "**A Management Architecture for a Multi-domain Content-Aware Network"**, TEMU Conf, Int'l Conf. on Telecommunications and Multimedia, 14-16 July, 2010, Crete, Greece.

[ALI-9] E.Borcoci, M.Stanciu, D.Niculescu (UPB), George Xilouris (DEM), "Quality of Services Assurance for Multimedia Flows based on Content-Aware Networking", Proceedings of the 4th International Conference on Communication Theory, Reliability, and Quality of Service (CTRQ 2011), Budapest Hungary, 19- 23 April, 2011. Best Paper award. http://www.iaria.org/conferences2011/CTRQ11.html

[ALI-10] E.Borcoci, M.Stanciu, D. Niculescu (UPB), D.Negru (LABRI), G.Xilouris (DEM), "Connectivity Services Management in Multi-domain Content-Aware Networks for Multimedia Applications", International Conference on Internet 2011, Luxembourg, June, 2011, http://www.ict-**ALICANTE**.eu/publication/2011/t6.1_infoware11.pdf

[ALI-11] Vorniotakis, N.; Xilouris, G.; Gardikis, G.; Zotos, N.; Kourtis, A. (DEM); Pallis, E.(TEIC); , **"A preliminary implementation of a Content-Aware network node,"** Multimedia and Expo (ICME), 2011 IEEE International Conference on , vol., no., pp.1-6, 11-15 July 2011, doi: 10.1109/ICME.2011.6011967

[ALI-12] D.Niculescu, M.Stanciu, M.Vochin, E.Borcoci, (UPB) N.Zotos (DEM), "Implementation of a Media Aware Network Element for Content Aware Networks", Proceedings of the 4th International Conference on Communication Theory, Reliability, and Quality of Service (CTRQ 2011), Budapest Hungary, 19- 23 April, 2011

[ALI-13] E.Borcoci (UPB), G.Carneiro (INESC), R.Iorga (UPB), **"Hybrid Multicast Management in a Content Aware Multidomain Network",**NetWare, 2011/AFIN **2011,** The Third International Conference on Advances in Future Internet, 21-27 August 2011, Nice, France**,** http://www.ict-**ALICANTE**.eu/publication/2011/t6.1_netware11.pdf, http://www.thinkmind.org/index.php?view=article&articleid=afin_2011_5_20_70107

[ALI-14] E.Borcoci, S.Obreja (UPB), C.Timmerer (UNI-KLU), G.Xilouris (DEM) "**Service Provider and Content Aware Network Provider Cross-Layer Optimisation of Multimedia Distribution**", ICME 2011, Workshop on Multimedia-Aware Networking 2011 (WoMAN 2011), Barcelona, Spain, July 11, 2011, http://woman2011.ict-**ALICANTE**.eu/

[ALI-15] R. Miruta, Badoi C., Borcoci E.(UPB) "New Packet Fields for Content Aware Classification", Proc. of AEECT 2011, Amman, Dec 2011, pp. 202-207 http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6132509 &url=http%3A%2F%2Fieeexplore.ieee.org %2Fxpls%2Fabs_all.jsp%3Farnumber%3D6132509

[ALI-16] M.Vochin, E.Borcoci, D.Niculescu,(UPB) ; et al. "**Scalability Analysis of a Media Aware Network Element**", Book Group Author(s): IEEE, Conf.: 20th European Signal Processing Conf. (EUSIPCO) Location: Bucharest, ROMANIA Date: AUG 27-31, 2012, Book Series: European Signal Proceedings Conference Pages: 2213-2217 Published: 2012.

[ALI-17] R.Miruta, C.Stanuica, E.Borcoci, (UPB) "Content Aware Classification Method", Book Group Author(s): IEEE, Conf.: 9th International Conference on Communications (COMM) Location: Bucharest, ROMANIA Date: JUN 21-23, 2012, Pages: 241-244 Published: 2012

[ALI-18] E.Borcoci, R.Miruta, S.Obreja,(UPB) **"Multi-domain Virtual Content-Aware Networks Mapping on Network Resources",** Book Group Author(s): IEEE, Conf.: 20th European Signal Processing Conf. (EUSIPCO) Location:

Bucharest, ROMANIA Date: AUG 27-31, 2012, Book Series: European Signal Proceedings Conference Pages: 2223-2227 Published: 2012.

[ALI-19] E.Borcoci, R.Miruta, S.Obreja, (UPB) „Inter-domain Peering in Content-Aware Networks for Multimedia Applications" , The Fifth International Conference on Communication Theory, Reliability, and Quality of Service CTRQ 2012, April 29 - May 4, 2012 - France http://www.iaria.org/conferences2012/CTRQ12.htm, ISBN: 978-1-61208-192-2 http://www.thinkmind.org/index.php?view=article&articleid=ctrq_2012_2_50_10097.

[ALI-20] E.Borcoci, R.Miruta, S.Obreja, (UPB), "Network Resources Allocation in Content-Aware Networks for Multimedia Applications",The Fourth International Conference on Evolving Internet, INTERNET 2012, June 24-29, 2012 - Venice, Italy, http://www.iaria.org/conferences2012/ProgramINTERNET12.html

[ALI-21] G. Gardikis, G. Xilouris, (DEM) "Network Cost Services for Network- Aware FI Applications" in FIArch Workshop on "Internet Architectural Research: Projects, Results and Gap Analysis", Position Paper, Brussels 27 Sept. 2012

[ALI-22] R.Iorga, E.Borcoci, R.Miruta, (UPB), "**Management Driven Multicast Protocol With End-to-End QoS**" –CTRQ 2013, The Sixth International Conference on Communication Theory, Reliability, and Quality of Service, CTRQ 2013, April 21 - 26, 2013 - Venice, Italy, ThinkMind // CTRQ 2013, The Sixth International Conference on Communication Theory, Reliability, and Quality of Service // View article ctrq_2013_1_20_10045

[ALI-23] R. Miruta, E.Borcoci, (UPB) E.Pallis, (TEIC) "Planning and provisioning of Virtual Content Aware Networks over IP infrastructures", , Telecommunications and Multimedia (TEMU), 2012 International Conference DOI: 10.1109/TEMU.2012.6294701, Publication Year: 2012 , Page(s): 118 – 123, IEEE Conference Publications

[ALI-24] G. Gardikis, K. Sarsembagieva, G. Xilouris, A. Kourtis, (DEM) **"An SNMP Agent for Active In-network Measurements",** in Proc. ICUMT 2012/IEEE, October 3-5, 2012, St. Petersburg, Russia

[ALI-25] G. Xilouris, G. Gardikis, K. Sarsembagieva, A. Kourtis, (DEM), M Sidibe (Viotech), D. Negru, Y. Chen (LABRI), "A Network Cost Provision Framework for Network-Aware Applications", submitted to Computer Networks

[ALI-26] R. Miruta, E.Borcoci, (UPB) „**Optimization of Overlay QoS Constrained Routing and Mapping Algorithm for Virtual Content Aware networks**", The Ninth International Conference on Networking and Services ICNS 2013, March 24 - 29, 2013 - Lisbon, Portugal, http://www.thinkmind.org/index.php?view=article&articleid=icns_2013_4_30_10174

[ALI-27] A. Pinto (INESC), **"Data plane efficiency of a hybrid multicast approach for content aware networks",**Future Network & Mobile Summit 2013 3 - 5 July 2013, Lisbon, Portugal

[ALI-28] R. Miruta, E. Borcoci (UPB), "Routing with QoS Constraints", UPB Scientific Bulletin, 2012, U.P.B. Sci. Bull., Series C, Vol. 75, Iss. 2, 2012, ISSN 2286-3540.

[ALI-29] P Anapliotis, E Markakis, A Sideris, E Pallis (TEIC), D Negru (LABRI), IC, **"A novel content-aware multipath routing concept exploiting random utility theory principles"** accepted, presented and published in the TEMU2012, Technically co-sponsored by IEEE Communication Society, http://www.researchgate.net/publication/233867750_A_novel_content-aware_multipath_routing_concept_exploiting_random_utility_theory_principles

[ALI-30] Sarsembagieva, K.; Gardikis, G.; Xilouris, G.; Kourtis, A. (DEM), **"A fast route planning algorithm for MPLS-TE,"** Telecommunications and Multimedia (TEMU), 2012 International Conference on , vol., no., pp.142,146, July 30 2012-Aug. 1 2012.

[ALI-31] G. Xilouris, G. Gardikis, K. Sarsembagieva, A. Kourtis (DEM), "SNMP-driven Active Measurements in DiffServ Networks," in Proc. IEEE ICC 2013, June 9-13, 2013, Budapest, Hungary (Accepted for Publication, to be Presented)

[ALI-32] K. Sarsembagieva, G. Gardikis, G. Xilouris, A. Kourtis, P. Demestichas (DEM), **"Efficient Planning of Virtual Network Services,"** in Proc. EUROCON IEEE 2013, July 1-4, 2013, Zagreb, Croatia (Accepted for Publication, to be Presented)

[ALI-33] V.Poenaru, E.Borcoci, M.Zotea, "Topology and Network Resources Discovery Protocol for Content-Aware Networks" INNOV 2013, November 17 - 22, 2013 - Lisbon, Portugal – **accepted.**

[ALI-34] E. Borcoci (UPB), D. Negru (LABRI), C.Timmerer (UNiklu) - Tutorial (3h) "Content-aware Networking: Future Internet Perspective", The Second International Conference on Advances in Advances in Future Internet , AFIN 2010, July 18 - 25, 2010 - Venice/Mestre, Italy http://www.iaria.org/conferences2010/TutorialsAFIN10.html

[ALI-35] E. Borcoci (UPB), D. Negru (LABRI) – Tutorial (3h), "Content Oriented Routing and Forwarding" NexComm Conference, April 29 - May 4, 2012 – France.http://www.iaria.org/conferences2012/TutorialsCTRQ12.html

[ALI-36] E.Borcoci, (UPB) Tutorial (2h) "New Networking Technologies Support for Media-oriented Applications", InfoSys 2013, March 24 - 29, 2013 - Lisbon, Portugal, http://www.iaria.org/conferences2013/TutorialsICNS13.html

[ALI-37] R.Miruta, "Content based processing networks", Ph.D. Thesis, Bucharest, UPB, March 2013.

[ALI-38] R.Iorga, "Hybrid Multicast Trees in Multimedia Environments", Ph.D. Thesis, Bucharest, UPB, March 2013.

Annex C: Distributed Framework for Edge and In-Network Media Adaptation

Abstract—**Existing and future media ecosystems need to cope with the ever-increasing heterogeneity of networks, devices, and user characteristics collectively referred to as (usage) context. The key to address this problem is media adaptation to various and dynamically changing contexts in order to provide a service quality that is regarded as satisfactory by the end user. The adaptation can be performed in many ways and at different locations, e.g., at the edge and within the network resulting in a substantial number of issues to be integrated within a media ecosystem. This paper describes research challenges, key innovations, target research outcomes, and achievements so far for edge and in-network media adaptation by introducing the concept of Scalable Video Coding (SVC) tunneling.**

Keywords-distributed adaptation decision-taking, SVC tunneling, research challenges, in-network adaptation, content-aware networking

I. Introduction & Problem Statement

The Universal Multimedia Access (UMA) [1] is omnipresent thanks to the evolution of device and network infrastructure technologies and we are now effectively entering the era of Universal Multimedia Experience (UME) [2]. An important aspect towards UMA and UME is the adoption of scalable media coding formats such as Scalable Video Coding (SVC) [3] enabling edge and in-network adaptation. This paper proposes the exploitation of these scalable media formats within the (core) network – with in-network adaptation enabled – in order to optimize the network resources utilization, and at the edge of the network, for the adaptation from/to heterogeneous formats, devices, and platforms. This is achieved by means of overlay networks, where the adaptation is coordinated in a distributed fashion. This innovation approach is referred to as SVC tunneling and the distributed coordination thereof aiming for a *better network resource utilization* while *maintaining a satisfactory Quality of Experience (QoE)*. The work is conducted as part of the EU FP7 Integrated Project ALICANTE (Media Ecosystem Deployment Through Ubiquitous Content-Aware Network Environments) [4] developing a media ecosystem comprising – among others – Home-Box (HB) and Content-Aware Network (CAN) overlay networks. For an overview of the ALICANTE architecture the reader is referred to [5].

The heterogeneity of devices, platforms, and networks is and most likely will be a constant companion within (future) media (Internet) ecosystems. Thus, we need to provide tools to cope with that heterogeneity in order to support a maximum of use cases while optimizing (network) resource utilization and improving QoE. One such tool is the SVC tunneling approach featuring edge and in-network media adaptation for which research challenges are highlighted in the following.

Distributed adaptation decision-taking framework:

Where to adapt? At the content source, within the network (with multiple options), at the receiving device, and combinations thereof.

When to adapt? At request and during the delivery enabling dynamic, adaptive streaming based on the users' context.

How often to adapt? Too often may increase the risk of flickering, whereas too seldom may result in stalling, both have a considerable impact on the QoE.

How to adapt? The optimization towards bitrate, resolution, framerate, signal-to-noise ratio (SNR), modality, accessibility, region-of-interest (ROI), etc. results in (too) many possibilities and often depends on the actual content, genre, and application.

Efficient, scalable SVC tunneling and signaling thereof:

Minimum *quality degradation* and *scalability* w.r.t. the number of parallel sessions and acceptable (end-to-end) latency. How can transcoding and adaptation steps be organized to minimize impact on QoS and video quality? How many parallel sessions can be supported on network and client equipment?

The impact on the QoS/QoE:

The *QoS/QoE trade-off* for the use cases and applications developed in ALICANTE. One example is the trade-off between quality degradation due to transcoding against the QoE gain of dynamic bitrate adaptation.

Possible *mappings* of QoS to QoE. Established network QoS parameters (such as packet loss, delay, and jitter) as well

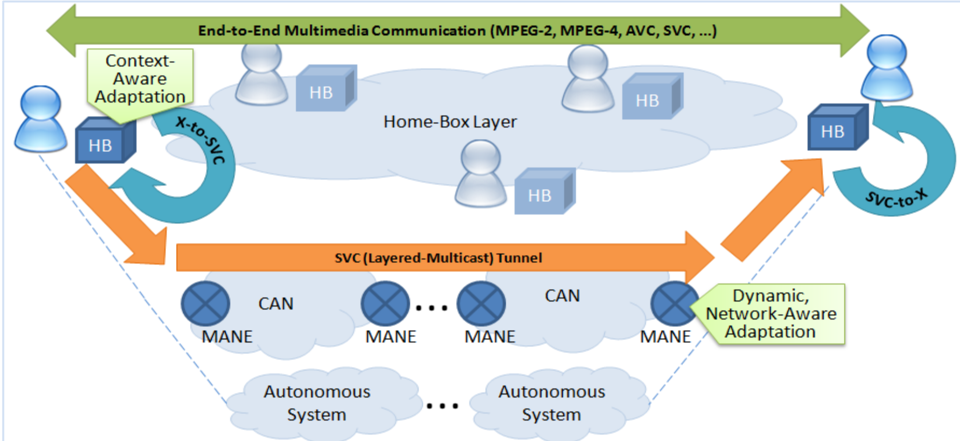


Figure C-1: Adaptation Framework Overview

as objective video quality are taken into account for estimating the viewing experience.

The adaptation framework and related key innovations are described in Section II. Target research outcomes, our test-bed setup, and scientific results so far are highlighted in Sections III and IV respectively. Section IV also presents test results comparing rate control modes for SVC tunneling. Finally, Section V concludes the paper and gives an outlook on future work.

II. The ALICANTE Adaptation Framework

A. Architecture

The ALICANTE system architecture introduces two new virtual layers, i.e., HB and CAN layers, on top of the existing network infrastructure. This approach brings both content-awareness to the network layer and context-awareness to the user environment.

A full description of the ALICANTE architecture can be found in [5]; this paper rather focuses on the adaptation framework of that architecture. Fig. C-1 shows an overview of the ALICANTE adaptation framework. Content delivery in the core network relies on scalable media formats such as SVC. This enables content-aware adaptation according to the network conditions at the CAN layer, i.e., within the Media-Aware Network Elements (MANEs).

Home-Boxes are enhanced home gateways with media processing capabilities. They can serve as home media servers, enable users to act as content providers, and keep track of the capabilities of connected terminals. Home-Boxes form a virtual HB layer that enables context-aware adaptation towards end user terminals and user preferences. For example, screen resolution and decoding capabilities are taken into account at content request time. For legacy terminals that do not support SVC, HBs are able to transcode content to non-scalable media formats (e.g., MPEG-2, MPEG-4 AVC). On the server side, corresponding HB layer entities are implemented as software modules.

B. Related Work

Similar to ALICANTE, several other research projects target media adaptation and content-aware networks. The FP7 Project ENVISION [6] proposes a multi-layered content distribution approach, targeting optimized end-to-end performance and content adaptation during distribution. However, it does not focus on QoE aspects on the client side. Dynamic and distributed adaptation of scalable multimedia content has been proposed by the FP6 Project DANAE [7]. With a focus on the MPEG-21 standard, it pioneered in the area of interoperable adaptation approaches. The FP6 Project ENTHRONE [8] developed a system architecture to cover the entire multimedia distribution chain, focusing on end-to-end QoS performance and network management. These projects tackle important aspects of media-aware adaptation along the delivery path. In the following we discuss several adaptation-related features of the ALICANTE architecture.

C. Adaptation Decision-Taking

Due to multiple locations within the delivery network where content may be subject to adaptation, we propose a distributed Adaptation Decision-Taking Framework (ADTF) that coordinates the local adaptation decisions of modules at the content source, the border to the user (Home-Box), and within the network at MANEs. Local adaptation decisions are taken based on an optimization algorithm, determining the most suitable adaptation for a given content, taking various aspects (e.g., user terminal capabilities, network monitoring results) into account. The various local adaptation decisions have different purposes, depending on the location they are performed in. For example, adaptation decisions in the network focus on dynamic adaptation towards network conditions, while adaptation decisions at the HB mainly target the capabilities of the user terminal and the QoE. The distribution of adaptation decisions also depends on the streaming mechanism, as, e.g., RTP multicast streaming is handled differently from HTTP streaming.

D. SVC Tunneling

In order to address the problem statement in Section I we propose an SVC (layered-multicast) tunnel, inspired by IPv6 over IPv4 tunnels. That is, within the CAN only scalable media resources – such as SVC – are delivered adopting a layered-multicast approach [9]. This allows the adaptation of scalable media resources by MANEs [10], implementing the concept of distributed adaptation [11], [12]. At the border to the user (Home-Box), adaptation modules are deployed enabling device-independent access to the SVC-encoded content by providing X-to-SVC and SVC-to-X transcoding/rewriting functions with X={MPEG-2, MPEG-4 Part 2, MPEG-4 Part 10 (AVC) etc.}. An advantage of this approach is the reduction of the load on the network (i.e., no duplicates), making it free for (other) data (e.g., more enhancement layers).

First measurements of SVC-based adaptation in an off-the-shelf WiFi router [13] have shown evidence to further pursue this kind of approach. More complex adaptation operations that are required to create scalable media resources, such as transcoding of media resources which have higher memory or CPU requirements, will be performed at the edge nodes, i.e., in the HB.

Note that SVC tunneling is also applicable to unicast scenarios due to dynamic SVC-based adaptation, although multicast scenarios bring higher gains in terms of network resource utilization.

E. In-Network Adaptation

MANEs perform dynamic in-network adaptation to mitigate the effects of network congestion. Each MANE has a local Adaptation Decision-Taking Engine (ADTE) that computes whether to adapt a media stream. The adaptation processes for multicast and unicast streaming have to be considered separately. Multicast streaming deploys RTP multi-session transmission (MST) mode, where SVC layers are transmitted over separate RTP sessions and are rearranged by the receiver. Thus, multicast trees for the different SVC layers are built. MANEs realize dynamic adaptation by pruning (or conversely grafting) the multicast tree corresponding to a specific SVC layer.

RTP-based unicast streaming is typically realized via single session transmission (SST) mode, where the entire SVC stream is packed into a single RTP session. In order to perform adaptation, a MANE de-packetizes the RTP stream, analyzes the SVC header, and filters SVC layers accordingly [14]. The RTP re-packetization module updates the sequence number field of the RTP packet headers if needed. Alternatively, unicast streaming could also be realized via MST mode, using separate ports for separate layers.

F. Scalability Considerations

The proposed techniques act on a per-flow basis, thus, some scalability considerations (in terms of number of concurrent flows) have to be taken into account. Adaptation decision-taking at a MANE has to handle many different flows in parallel, requiring a very lean and efficient implementation of the ADTE. The processing overhead can be controlled by the update frequency of adaptation decisions. For example, the decision to drop an SVC layer shall be triggered immediately when network monitoring indicates congestion, but the decision to add a layer back to the stream can be delayed by a scheduler until CPU utilization has declined to a certain threshold. In contrast to the MANE, adaptation decision-taking at the Home-Box has to take more parameters into account, including terminal capabilities and user preferences, but has fewer flows to handle. A Home-Box in a typical household might have to handle up to five concurrent flows. However, any adaptation or transcoding operations have way higher computational complexity and resource demands than the adaptation decision-taking.

Transcoding at the server side and at the Home-Box are computationally expensive parts of SVC tunneling. Transcoding to SVC on the server-side has only to be performed once per video and can be performed offline prior to streaming. Transcoding from SVC to other formats on the Home-Box demands less resources but the Home-Box has to be dimensioned to support a handful of concurrent flows.

In-network adaptation in MST mode relies on receiver-driven layered multicast, thus, the usage of SVC does not put any overhead on this approach. In SST mode, RTP de-packetization and re-packetization limit the number of concurrent flows. A prototype implementation on an off-the-shelf WiFi router supported concurrent adaptation of several flows in 2008 [13], dedicated hardware and improved algorithms may lead to a higher number of possible concurrent flows.

G. The ALICANTE Adaptation Framework Key Innovations

The corresponding key innovations are summarized as follows:

- Better network *resource utilization* based on adaptation and *maintaining a satisfactory QoE/QoS*: Content is encoded in or transcoded to scalable media formats such as SVC for efficient layered multicast distribution enabling in-network adaptation. End users and network devices provide

Table C-1: Y-PSNR results of SVC layers for the hall_monitor sequence with various encoders and rate control modes.

Target Quality	bSoft (VBR)					MainConcept (VBR)					MainConcept (CBR)				
	bitrate [kbps]	*L3 [dB]*	*L2 [dB]*	*L1 [dB]*	*L0 [dB]*	*bitrate [kbps]*	*L3 [dB]*	*L2 [dB]*	*L1 [dB]*	*L0 [dB]*	*bitrate [kbps]*	*L3 [dB]*	*L2 [dB]*	*L1 [dB]*	*L0 [dB]*
Q1	4482	44.74	33.05	26.89	23.20	3270	42.97	39.15	35.94	32.96	3095	43.87	42.64	41.03	37.74
Q2	2446	42.03	32.95	26.89	23.20	1867	40.10	36.62	33.38	30.39	2202	42.30	41.07	39.53	36.30
Q3	1244	39.84	32.77	26.87	23.21	1191	37.79	34.22	30.88	27.84	1622	40.96	39.75	38.32	35.09
Q4	699	37.83	32.48	26.86	23.23	816	35.46	31.74	28.34	25.30	1058	38.68	37.43	36.03	32.62

QoE/QoS feedback to the ADTF, to adjust the service in a distributed and dynamic way.

- Context information from multiple receivers is aggregated at MANEs and used for *local* adaptation decision-taking. Additionally, adaptation decisions are propagated within the media delivery network enabling *distributed* adaptation decision-taking. For example, upstream to the server in case of RTP-like streaming or downstream to the receiver in case of HTTP-like streaming.
- *Distributed coordination* for optimal adaptation and improved bandwidth usage involves the active participation of multiple entities across the media delivery network such as adaptation decision-taking, actual adaptation, and QoE/QoS probes.

III. Targeted Research Outcomes and Test-Bed

A. Targeted Research Outcomes

Based on the ALICANTE architecture and the aforementioned key innovations, we target the following research outcomes:

- Guidelines for *scalable media encoding/transcoding parameters* (with SVC as example) for a given set of use cases enabling in-network adaptation. They will answer questions like how many SVC layers at which bitrate, framerate, and SNR are the optimal choice for the use cases developed in ALICANTE.
- Guidelines for the *distributed adaptation decision-taking framework* which basically provides answers to the open questions raised in Section I, i.e., where to adapt, when to adapt, and how (often) to adapt.
- Enhancement of the decision-taking algorithm through the exploitation of the active and passive monitoring framework based on flow- and content-awareness at MANEs.
- Enhancement of the SVC adaptation based on network load/conditions and QoS constraints using a content-aware approach.
- Assessment of the performance and scalability (e.g., number of flows, flow traffic profile) of the developed in-network adaptation mechanisms, in terms of computing resources utilized (e.g., CPU and memory) as well as network related metrics (e.g., processing delay per flow, maximum achieved bandwidth).
- Based on our test-bed and pilots we will determine the (end-to-end) *delay* (taking real-time applications into account), minimum *quality degradation* (measured via objective methods, i.e., PSNR, and subjective methods, i.e., MOS), and *scalability* w.r.t. the number of parallel sessions which should not increase exponentially.
- Mappings of network and device monitoring parameters that enable the prediction of the QoE and the validation thereof through subjective quality assessments.
- A holistic approach for in-network adaptation applying different adaptation policies per content-aware virtual network.

B. Proposed Test-bed Setup

The proposed test-bed setup for our evaluations is described below and depicted in Fig. C-2.

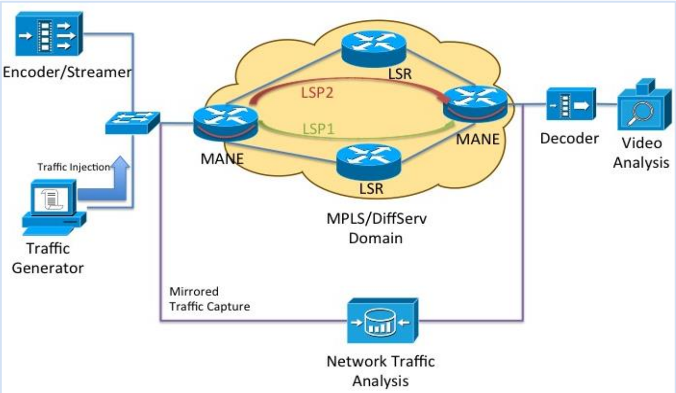


Figure C-2 : Test-bed Setup

The content is encoded or transcoded into SVC by the Encoder/Streamer module. A Traffic Generator creates cross-traffic that is used for the evaluation of DiffServ and adaptation functionalities of MANEs. The mixed traffic is ingested through a MANE into the core network and is simultaneously captured by a Network Traffic Analysis tool. That ingress MANE performs dynamic SVC adaptation and content-aware forwarding over label-switched paths (LSPs). The traffic is routed through the network via Label Switching Routers (LSRs). At the egress point of the network, we capture the traffic again for network traffic analysis. The SVC content is decoded or transcoded to a suitable terminal format by the Decoder module. The Encoder/Streamer module and the Decoder module correspond to HBs in the ALICANTE architecture and also form the ingress and egress points for SVC tunneling. Finally, the received content is analyzed by the Video Analysis module. The video quality analysis comprises objective methods and a QoS-QoE mapping model [14]. This mapping will be validated using subjective tests.

This test-bed integrates our previous setups of [14], [15], and [16].

IV. Scientific Results Achieved

The research challenges and open issues including a description of use cases in question are further detailed in [14]. In this section we summarize scientific results and present test results for comparing rate control modes for SVC tunneling.

A. Achieved Results

The quality impact of SVC tunneling is investigated in [15] using MPEG-2 as starting point and providing a baseline for further research. The results indicated a total decrease in luminance-component PSNR (Y-PSNR) of 2.1 dB, with around 1/3 of the quality impact attributed to the initial MPEG-2-to-SVC transcoding step. A bitrate increase of 43% is required to compensate the quality loss which is still less than the necessary bandwidth for MPEG-2 simulcast-based streaming. Based on the test-bed setup described in [15], we performed further tests for comparing the impact of rate control modes for SVC tunneling as detailed later on.

An initial performance evaluation of SVC streaming and real-time in-network adaptation is reported in [16]. Further evaluations with enabled end-to-end QoS control are presented in [14] including a model for QoS-QoE mapping. The video streaming system introduces a cross-layer QoS mapping based framework for media- and user/terminal-aware transmission and management. The experimental results indicate the advantage of such an adaptation system that facilitates the control of bandwidth utilization to obtain an improved perceived video quality for end users.

Finally, initial subjective quality assessments for an application that may benefit from ALICANTE's media ecosystem are described in [17]. In the future it is anticipated to apply such subjective tests on the work conducted in [14], [15], and [16] respectively. Furthermore, we are currently working on combining [15] with the QoS-QoE mapping model of [14] and the deployment of multi-video rate allocation for SVC tunneling.

B. Comparing rate control modes for SVC Tunneling

We extended the aforementioned tests of [15] for comparing SVC tunneling of variable bitrate (VBR) encoding mode against constant bitrate (CBR) encoding mode using the following setup. The test was performed with the test sequences *foreman*, *container*, *hall_monitor*, and *stefan*, each having a resolution of 352x288 and 25 fps framerate. The test sequences were initially encoded to MPEG-2, transcoded in a first transcoding step to SVC using pixel-domain transcoding (PDT), and in a second transcoding step back to MPEG-2 using PDT. These transcoding scenarios were performed for VBR and CBR encodings separately. For comparing the required bandwidth of SVC tunneling with MPEG-2 simulcast, we selected for each extracted SVC layer an MPEG-2 encoding with best matching Y-PSNR.

Table C-2: Bjontegaard Delta for SVC tunneling.

Sequence	bSoft (VBR)		MainConcept (VBR)		MainConcept (CBR)	
	BD-PSNR [dB]	*BD-bitrate [%]*	*BD-PSNR [dB]*	*BD-bitrate [%]*	*BD-PSNR [dB]*	*BD-bitrate [%]*
foreman	-2.08	50.3	-2.03	53.7	-2.40	61.6
container	-1.57	38.2	-1.99	51.0	-2.91	66.9
hall_monitor	-0.75	22.6	-1.40	54.1	-1.82	73.6
stefan	-2.59	41.0	-2.09	32.1	-2.88	53.4
Average	**-1.74**	**38.04**	**-1.88**	**47.7**	**-2.50**	**63.9**

One challenge in this setup is the selection of a suitable quantization parameter (QP) or target bitrate for the SVC encoding in the first transcoding step. In [15], the same target bitrate was used for the initial MPEG-2 encoding and the SVC encoding. Such configuration is not flexible when it comes to saving network resources in streaming scenarios. Thus, we chose an experimental approach were the original sequence is encoded to SVC with several target qualities (i.e., QP or target bitrate) and then the configuration that yields a Y-PSNR just above that of the MPEG-2 encoding is selected. For the second transcoding step (back to MPEG-2), we applied again the target quality of the initial MPEG-2 encoding.

The SVC encoding was configured with four medium-grained scalability (MGS) layers. We tested two industry-grade SVC encoders, i.e., MainConcept v1.5 [18] and bSoft v120403 [19]. MPEG-2 encoding was performed via FFmpeg v0.8 [20]. The bSoft encoder distributes transform coefficients to create MGS enhancement layers. On the other hand, the MainConcept encoder performs re-quantization to obtain those MGS layers. Compared to the highest layer, we reduced the QP by 6 per MGS layer for VBR or conversely the target bitrate by 30% for CBR. Since the bSoft encoder always yielded better rate-distortion (RD) performance for VBR mode, we did not perform CBR mode tests for the bSoft encoder.

The starting points of the test are four SVC encoding configurations (*Q1*, *Q2*, *Q3*, *Q4*) with highest layer QP of {28, 24, 20, 16} for VBR and target bitrate of {1, 1.5, 2, 3} Mbps for CBR. The qualities of the SVC layers (labeled *L3* for highest layer and *L2*, *L1*, *L0* for the lower layers respectively) of the *hall_monitor* sequence are exemplarily shown in Table C-1. While the bSoft encoder yields good overall RD performance, the automatic distribution of transform coefficients allocates little quality to the lower layers (due to a uniform rate distribution among layers) compared to our configuration of the MainConcept encoder.

Table C-3: Comparison of required bandwidths for SVC tunneling vs. MPEG-2 simulcast.

Target Quality	bSoft (VBR)		MainConcept (VBR)		MainConcept (CBR)	
	SVC tunneling [kbps]	*MPEG-2 simulcast [kbps]*	*SVC tunneling [kbps]*	*MPEG-2 simulcast [kbps]*	*SVC tunneling [kbps]*	*MPEG-2 simulcast [kbps]*
Q1	5333	3041	3694	3454	3286	4721
Q2	3446	2025	2418	2082	2242	3191
Q3	2201	1452	1650	1277	1687	2093
Q4	1438	1102	1132	900	1109	1287
Average	**3105**	**1905**	**2224**	**1928**	**2081**	**2823**

The Bjontegaard Delta (BD) [21] results for the two transcoding steps are shown in Table C-2. The BD is measured between the initial and final MPEG-2 encoding. As mentioned before, we applied a flexible approach for the target quality of SVC encoding. Thus, the BD is applicable neither to the MPEG-2-to-SVC transcoding step nor the SVC-to-MPEG-2 transcoding step, but only to the result of the entire transcoding chain.

Sequences with lower spatial detail and lower amount of movement (such as *hall_monitor*, *container*) typically show less quality degradation than those with higher amounts. The overall results show lower quality impact for VBR mode (-1.74 dB for bSoft encoder, -1.88 dB for MainConcept encoder on average) than for CBR mode (-2.50 dB on average).

Table C-3 presents the comparison of average required bandwidths for SVC tunneling and MPEG-2 simulcast streaming. Columns labeled *SVC tunneling* show required bandwidths for delivering the content which has been transcoded from MPEG-2 to SVC (i.e., first transcoding step). For the second transcoding step, the content is transcoded back into the final MPEG-2 encoding. The required bandwidths for MPEG-2 simulcast (of the initial MPEG-2 encoding) with the same quality (in terms of Y-PSNR) as that final MPEG-2 encoding are shown in columns labeled *MPEG-2 simulcast*.

Only CBR mode yields lower overall bandwidth requirements for full SVC tunneling than for equivalent MPEG-2 simulcast, reducing the required bandwidth by up to 32% (and 26% on average). For the tested configurations, SVC tunneling with VBR mode performs worse than equivalent MPEG-2 simulcast, even though it yields less quality degradation. This is attributed to the comparatively high quality of lower SVC layers in CBR mode (cf. Table C-1), which manifests in higher bitrates of MPEG-2 simulcast in order to match that quality. We argue that the bandwidth efficiency of SVC tunneling depends on the configuration of lower SVC layers rather than on the encoder implementation. Furthermore, the number of SVC enhancement layers (ELs) plays an important role for the bandwidth efficiency of SVC tunneling. Future work will investigate whether the additional MPEG-2 encodings for simulcast outweigh the SVC coding overhead of additional ELs in terms of required bandwidth. Note that SVC tunneling with VBR mode may still be favorable to MPEG-2 simulcast in scenarios where only one of the two transcoding steps is needed (e.g., if the client's media player supports SVC), since every transcoding step has an impact on video quality.

V. Conclusions

In this paper we have presented research challenges and key innovations for edge and in-network adaptation featuring a distributed ADTF and SVC tunneling. Towards targeted research outcomes in the FP7 ALICANTE project, we proposed an integrated test-bed. Besides presenting a selection of scientific results achieved so far, this paper provides further results in SVC tunneling with focus on comparing the impact of VBR and CBR encoding modes with respect to quality degradation and bandwidth efficiency. The results indicate smaller quality impact for VBR mode (-1.74 dB and -1.88 dB, depending on the encoder) than for CBR (-2.50 dB), but the comparison of required bandwidth only yields bandwidth reduction (of 26%) for SVC tunneling with CBR mode. More details and a complete description can be found in [22]

Future work targets high-definition content, subjective tests, and integrating the QoS-QoE mapping of [14] and multi-video rate allocation into the proposed test-bed.

Acknowledgment

This work was supported in part by the EC in the context of the ALICANTE project (FP7-ICT-248652). The authors thank bSoft Ltd. and MainConcept GmbH for providing their encoders.

References

[1] R. Mohan, J. R. Smith, and C.-S. Li, "Adapting multimedia Internet content for universal access," IEEE Transactions on Multimedia, vol.1, no.1, pp.104-114, Mar. 1999.

[2] F. Pereira and I. Burnett, "Universal multimedia experiences for tomorrow," IEEE Signal Processing Magazine, vol.20, no.2, pp. 63- 73, Mar. 2003.

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," IEEE Transactions on CSVT, vol.17, no.9, pp.1103-1120, Sept. 2007

[4] ALICANTE web site, http://ict-alicante.eu/. June 2012.

[5] E. Borcoci, D. Negru, and C. Timmerer, "A Novel Architecture for Multimedia Distribution Based on Content-Aware Networking," Proc. CTRQ'10, Athens/Glyfada, Greece, June 2010.

[6] ENVISION web site, http://envision-project.org/. June 2012.

[7] European R&D Projects: DANAE, web site, http://cordis.europa.eu/projects/rcn/71233_en.html. June 2012.

[8] ENTHRONE web site, http://enthrone.org/. June 2012.

[9] S. McCanne, V. Jacobson, and M. Vetterli, "Receiver-driven layered multicast," Proc. of SIGCOMM'96, pp.117-130, Aug. 1996.

[10] S. Wenger, Y.-K. Wang, and T. Shierl, "Transport and Signaling of SVC in IP Networks," IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 9, pp. 1164–1173, Sept. 2007.

[11] S. Devillers, C. Timmerer, J. Heuer, and H. Hellwagner, "Bitstream syntax description-based adaptation in streaming and constrained environments," IEEE Transactions on Multimedia, vol.7, no.3, pp. 463- 470, June 2005.

[12] J. Chakareski, "In-Network Packet Scheduling and Rate Allocation: A Content Delivery Perspective," IEEE Transactions on Multimedia, vol.13, no.5, pp.1092-1102, Oct. 2011.

[13] I. Kofler, M. Prangl, R. Kuschnig, and H. Hellwagner, "An H.264/SVC-based adaptation proxy on a WiFi router," Proc. NOSSDAV'08, Braunschweig, Germany, 2008.

[14] B. Shao et al., "An Adaptive System for Real-Time Scalable Video Streaming with End-to-End QoS Control," Proc. WIAMIS'10, Desenzano Del Garda, Italy, Apr. 2010.

[15] M. Grafl, C. Timmerer, and H. Hellwagner, "Quality Impact of Scalable Video Coding Tunneling for Media-Aware Content Delivery," Proc. ICME'11, Barcelona, Spain, July 2011.

[16] N. Zotos, G. Xilouris, D. Renzi, B. Shao, and A. Kourits, "Performance evaluation of H264/SVC streaming system featuring real-time in-network adaptation," Proc. IWQoS'11, San Jose, California, June 2011.

[17] M. Waltl, C. Timmerer, and H. Hellwagner, "Improving the Quality of Multimedia Experience through Sensory Effects," Proc. QoMEX'10, Trondheim, Norway, June 2010.

[18] MainConcept web site, http://mainconcept.com/. May 2012.

[19] bSoft web site, http://bsoft.net/. May 2012.

[20] FFmpeg web site, http://ffmpeg.org/. May 2012.

[21] G. Bjontegaard, "Improvements of the BD-PSNR model," ITU-T Q.6/SG 16 Video Coding Experts Group, doc. VCEG-AI11, Berlin, Germany, 2008.

[22] Michael Grafl, Christian Timmerer, Hermann Hellwagner, George Xilouris, Georgios Gardikis, Daniele Renzi, Stefano Battista, Eugen Borcoci, and Daniel Negru, "Scalable Media Coding enabling Content-Aware Networking", IEEE MultiMedia, IEEE Computer Society, vol. 20, no. 2, pp. 30-41, June 2013.

Curriculum Vitae

Daniel Négru

Date de Naissance: 24/05/1979
Grade : Maitre de Conférences (depuis 2007)
Etablissement d'affectation : Institut Polytechnique de Bordeaux (IPB)
Section CNU : 27
Unité de recherche d'appartenance : UMR 5800, LaBRI
Equipe de recherche : PROGRESS (Thème Comet)
PES : Oui (A, depuis 2010).

Tel.: 05.40.00.37.97
e-mail : daniel.negru@labri.fr

Résumé des activités de recherche :
Les deux grands thèmes de recherche sur lesquels se basent mes travaux sont les suivants:

1. Convergence des réseaux et services de prochaine génération, leur implication au sein des réseaux domestiques (Home Networking) et extension (Extended-Home) pour les utilisateurs finaux, en termes de services haut niveau, adaptation, qualité et mobilité.
2. Evolution des réseaux de médias au sein de l'Internet du Futur, vers une approche consciente du contenu et du contexte (Content- and Context-Aware Network) et des applications conscientes des caractéristiques réseaux (Network-Aware Applications).

L'originalité des recherches menées s'inscrit autour de la mise en relation de plans virtuels à différents niveaux : un overlay de services et un overlay de réseaux communiquent et s'échangent des informations afin de mieux effectuer leurs tâches respectives et ainsi être plus efficaces. Les solutions proposées permettent de lever des verrous liés à la scalabilité, d'une part, et la performance, d'autre part, d'une telle approche Contenu-Consciente (Content-Aware) du réseau. De la même manière, au niveau utilisateur, les solutions proposées mettent en avant une approche contextualisante et adaptative facilitant la relation aux services haut niveau. Afin de démontrer l'efficience de certains aspects, la modélisation et simulation ne suffisent pas, il devient nécessaire de concevoir, évaluer et valider des systèmes complets, englobant l'électronique et le middleware, en addition aux protocoles et algorithmes.
Les résultats marquants des travaux effectués se retrouvent dans différentes publications, au travers de plusieurs projets de recherche collaboratifs et ont abouti à la réalisation de deux principaux prototypes d'équipements, d'une part, et de nombreux logiciels, d'autre part, en collaboration avec des industriels. La standardisation des solutions proposées est également un objectif, pour lequel les premiers travaux ont abouti. De manière plus précise pour chacun des thèmes abordés:

Thème 1 (Convergence des réseaux et services) *:*

- *Principales publications* : [3], [6], [8], [12], [14], [19], [21], [22], [23], [25], [27], [39] ;
- *Projets* : ALICANTE, ARDMAHN, MultiBox, ConvergenceIMS ;
- *Prototypes* : Home-Box (équipement), A-PSQA QoE tool, ConvergenceIMS adaptive gateway, DASH-VLC encoder-decoder-streamer, DASH proxy gateway;
- *Groupes de Standardisation* : ETSI, IETF, HGI (à venir);

Thème 2 (Evolution des réseaux de médias) :

- *Principales publications* : [1], [2], [7], [9], [16], [24], [30], [32], [33], [35], [38] ;
- *Projets* : ALICANTE, ENTHRONE, ADAMANTIUM, SEA, DISEDAN (à venir), T-NOVA (à venir) ;
- *Prototypes* : MANE (CAN Router : équipement), Adaptive IMS Management System, SVC Codec ;
- *Groupes de Standardisation* : FIA, BroadBand Forum, MPEG.

1. Publications et production scientifique

Tableau Récapitulatif

	Bon & Très bon[1] (en gras ci-dessous)	Moyen[7]
Chapitres d'ouvrages	**3**	2
Articles de revues/journaux/ magazines	**5**	4
Articles de conférences	**17**	12
Tutoriaux	**0**	3
Documents de standardisation	**4**	1
Articles en soumission	**1 journal 1 conférence**	-
Total	**31**	22

Chapitres d'ouvrages internationaux (avec comité de lecture)

[1] [DNSTI11]
M. Grafl, C. Timmerer, H. Hellwagner, D. Negru, E. Borcoci, D. Renzi, A. Mevel, A. Chernilov, "Scalable Video Coding in Content-Aware Networks: Research Challenges and Open Issues", Springer Trustworthy Internet Book, 2011.

[2] [DNFIA11]
H. Koumaras, D. Négru et al., "Media Ecosystems: A Novel Approach for Content-Awareness in Future Networks", Future Internet: Achievements and Promising Technology, Springer Verlag, pp.369-380. May 2011.

[3] [DNRPB09]
D. Négru, Y. Hadjadj-Aoul, A. Mehaoua, A. Kourtis, "IPv6 over DVB-T: Mobility Issues, Challenges, and Enhancements", River Publishers Book on Mobility Management and Quality-of-Service for Heterogeneous Networks, 2009.

[4] [DNLNCS04]
D. Négru, A. Mehaoua, A. Kourtis, E. Lebars, "Large-Scale Mobile Multimedia Service Management in Next Generation Networks", Springer Lecture Notes in Computer Science on Management of Multimedia Networks and Services, 2004.

---

[7] Classement personnel où le niveau des contributions est établi en fonction de « impact factor » pour les journaux, du taux d'acceptation pour les conférences (Bon & Très bon < 40%, Moyen > 40%), ainsi que de la notoriété perçue au sein de la communauté (valeur subjective utilisée quand les éléments précédents ne sont pas disponibles).

[5] [DNLNCS03]
Y. Hadjadj-Aoul, D. Négru, A. Nafaa, A. Mehaoua, “M3G: A Mobile Multicast Multimedia Gateway for Seamless IPv4/IPv6 Transition”, Springer Lecture Notes in Computer Science on Management of Multimedia Networks and Services, 2003.

Articles de revues / journaux internationaux (avec comité de lecture):

En soumission :

[6] [DNEJO13]
J. MongayBatalla, A. Bęben, P. Krawiec, D. Négru, “EMO algorithm for two-phase selection process in Content Aware Networks”, under review in Special Issue of the European Journal of Operational Research on Evolutionary Multiobjective Optimization, September 2013.

Publiés (ou en cours de publication) :

[7] [DNIMM13]
M. Grafl, C. Timmerer, H. Hellwagner, G. Xilouris, G. Gardikis, D. Renzi, S. Battista, E. Borcoci, D. Negru, “Scalable Media Coding enabling Content-Aware Networking”, IEEE MultiMedia, Volume: 20, Issue: 2, pp. 30-41, April-June 2013.

[8] [DNEJS13]
F. Duhem, F. Muller, W. Aubry, B. Le Gal, Daniel Négru, P.Lorenzini, “Design Space Exploration for Partially Reconfigurable Architectures in Real-Time Systems”, Elsevier Journal of Systems Architecture, Volume 59, Issue 8, pp. 571–581 2012, September 2013.

[9] [DNICM12]
G. Gardikis, G. Xilouris, T. Kourtis, E. Pallis, M. Sibidé, D. Négru, “Cross-layer Monitoring in IPTV Networks”, IEEE Communication Magazine, Vol. 50, Issue 7, July 2012.

[10] [DNIJDM11]
G. Gardikis, G. Xilouris, M.-J. Montpetit, A. Vanelli-Coralli, D. Negru, “IP and Broadcasting Systems Convergence”, Hindawi Int. Journal of Digital Multimedia Broadcasting, 2011.

[11] [DNIJDMB11]
E. Pallis, D. Negru, A. Bourdena, “Hybrid Terrestrial-Satellite DVB/IP Infrastructure in Overlay Constellations for Triple-Play Services Access in Rural Areas”, Hindawi Int. Journal of Digital Multimedia Broadcasting, 2011.

[12] [DNSMT10]
J. Arnaud, D. Négru, M. Sidibé, J. Pauty, H. Koumaras, “Adaptive IPTV services based on a novel IP Multimedia Subsystem”, Springer Multimedia Tools and Applications Journal, 2010.

[13] [DNIJCE08]
H. Koumaras, D. Négru, F. Liberal, J. Arauz, A. Kourtis, “ADAMANTIUM: Enhancing IMS with a PQoS-aware Multimedia Content Management System”, International Journal of Control Engineering and Applied Informatics 10, Volume 2, 2008.

[14] [DNECC06]
D. Négru, A. Mehaoua, Y. Hadjadj-Aoul, C. Berthelot, “Dynamic Bandwidth Allocation for Efficient support of Concurrent Digital TV and IP Multicast Services in DVB-T Networks”, Elsevier Computer Communications Journal 29, 2006.

[15] [DNWSEA04]
D. Négru, G. Xilouris, Y. Hadjadj-Aoul, A. Mehaoua, E. Pallis, "Mobility Issues in a DVB-T Environment", WSEAS Transactions on Communications Volume 3, Issue 2, 2004.

Articles dans actes de conférences internationales (avec comité de lecture):

En soumission :

[16] [DNPC14]
S. Ait Chellouche, D. Négru, J. Mongay Batalla, "Context-aware multimedia services provisioning in Future Internet using ontology and rules", under review in IEEE PerCom 2014, September 2013.

Publiés (ou en cours de publication) :

[17] [DNISCC13]
M. Grafl, C. Timmerer, H. Hellwagner, W. Cherif, D. Negru, S. Battista, "Scalable Video Coding Guidelines and Performance Evaluations for Adaptive Media Delivery of High Definition Content", IEEE Symposium on Computers and Communication (ISCC 2013), Split, Croatia, July 2013.

[18] [DNTEM12]
W. Aubry, D. Négru, B. Le Gal, D. Dallet, "Spatial Downsizing impact in the Transrating tradeoff for content/context awareness in media network", in proc. of the IEEE International Conference on Telecommunications and Multimedia (TEMU 2012), Heraklion, Crete, Greece, July 2012.

[19] [DNTEMU12]
P. Anapliotis, E. Markakis, A. Sideris, E. Pallis, D. Négru, "A novel content-aware multipath routing concept exploiting random utility theory principles", in proc. of the IEEE International Conference on Telecommunications and Multimedia (TEMU 2012), Heraklion, Crete, Greece, July 2012.

[20] [DNEUSIP12]
W. Aubry, D. Négru, B. Le Gal, S. Desfarges, D. Dallet, "A Generic Video Adaptation Framework Towards Content - and Context - Awareness in Future Networks", in proc. of the 20th IEEE European Signal Processing Conference (EUSIPCO 2012), Bucharest, Romania, August 2012.

[21] [DNIWCMC12]
J. Arnaud, D. Négru,"Flow scheduling at multihomed devices during vertical handover in NGN", in proc. of the 8th IEEE International Wireless Communications and Mobile Computing Conference (IWCMC 2012), Cyprus, August 2012.

[22] [DNIESCC12]
J. Arnaud, D. Négru, "Mobility-Aware P/I-CSCF: a solution for achieving seamless handover in IMS", in proc. of the IEEE Symposium on Computers and Communications (ISCC'12), Cappadocia, Turkey, July 2012.

[23] [DNISCC12]
S. AitChellouche, D. Négru, Y. Chen,"Home-Box-assisted Content Delivery Network for Internet Video-on-Demand Services", in proc. of the IEEE Symposium on Computers and Communications (ISCC'12), Cappadocia, Turkey, July 2012.

[24] [DNICC12]
W. Cherif, A. Ksentini, M. Sidibé, D. Négru,"A_PSQA: PESQ-like non-intrusive tool for QoE prediction in VoIP Services", in proc. of the International Conference on Communications (ICC 2012), Ottawa, Canada, June 2012.

[25] [DNDAIS12]
P. Rodrigues, D. Bromberg, L. Réveillère, D. Négru,"ZigZag: a Middleware for Service Discovery in Future Internet", in proc. of the 7th International Federated Conference on Distributed Computing Techniques, DAIS conference (DisCoTec'12), Stockholm, Sweden, June 2012.

[26] [DNICE11]
W. Aubry, B. Le Gal, D. Dallet, S. Desfarges, D.Négru,"A system approach for reducing power consumption of multimedia devices with a low QoE impact", in proc. of the IEEE International Conference on Electronics, Circuits and Systems (ICECS 2011), Beirut, Lebanon, December 2011.

[27] [DNMID11]
P.Rodrigues, L.Reveillere, D. Bromberg, D. Négru, "Scalable and Interoperable Service Discovery for Future Internet", in proc. of the ACM/IFIP/USENIX 12th International Middleware Conference, M-MPAC Workshop, Lisbon, Portugal, December 2011.

[28] [DNICME11]
W. Cherif, A. Ksentini, M. Sidibé, D. Négru,"A_PSQA: Efficient Real-Time Video Streaming QoE Tool in a Future Media Internet Context", in proc. of the 2011 IEEE International Conference on Multimedia and Expo (ICME 2011), Barcelona, Spain, July 2011.

[29] [DNWoM11]
E. Pallis, V. Zacharopoulos, D. Negru, P. Anapliotis, "Enhancing Legacy Infrastructures with Content-Aware Enablers Towards a Networked-Media Platform", in proc. of the 2011 IEEE International Conference on Multimedia and Expo (ICME 2011), WoMAN Workshop, Barcelona, Spain, July 2011.

[30] [DNWoMA11]
S. Ait Chellouche, D. Négru, Y. Chen, W. Aubry, "Home-Boxes Support for an Efficient Video-on-Demand Distribution", in proc. of the 2011 IEEE International Conference on Multimedia and Expo (ICME 2011), WoMAN Workshop, Barcelona, Spain, July 2011.

[31] [DNISCC11]
G. Gardikis, G. Xilouris, A. Kourtis, D. Negru, Y. Chen, P. Anapliotis, E. Pallis,"Media Ecosystem Deployment in a Content-Aware Future Internet Architecture", in proc. of the IEEE Symposium on Computers and Communications (ISCC'11), Kerkyra, Greece, June 2011.

[32] [DNINT11]
E. Borcoci, M. Stanciu, D. Niculescu, D. Negru, G. Xilouris, "Connectivity Services Management in Multi-domain Content-Aware Networks for Multimedia Applications", in proc. of the Third International Conference on Evolving Internet (INTERNET 2011), Luxembourg, June 2011.

[33] [DNCCNC11]
S. Ait Chellouche, D. Negru, E. Borcoci, E. Lebars,"Content-Aware Distributed Multimedia Provisioning based on Anycast Model towards Future Media Internet", in proc. of the 2011

IEEE Consumer Communications & Networking Conference (CCNC'11), Las Vegas, Nevada, USA, January 2011.

[34] [DNGLOB10]
S. Ait Chellouche, D. Négru, E. Lebars, E. Borcoci,"Anycast-based Context-Aware Server Selection Strategy for VoD Services", in proc. of the 2010 IEEE Global Communications Conference (GLOBECOM'10) Workshops, Miami, Florida, USA, December 2010.

[35] [DNCTRQ10]
E. Borcoci, D. Negru, C. Timmerer, "A Novel Architecture for Multimedia Distribution based on Content-Aware Networking", in proc. of the Third International Conference on Communication Theory, Reliability, and Quality of Service (CTRQ 2010), Athens, Greece, June 2010.

[36] [DNCCNC10]
S. Ait Chellouche, J. Arnaud, D. Negru,"Flexible User Profile Management for Context-Aware Ubiquitous Environments", in proc. of the 2011 IEEE Consumer Communications & Networking Conference (CCNC'10), Las Vegas, Nevada, USA, January 2010.

[37] [DNMOBI09]
H. Koumaras, J. Arnaud, D. Négru, A. Kourtis, "An Experimental Approach of Video Quality Level Dependence on Video Content Dynamics", in proc. of the 5th ICST International Mobile Multimedia Communications Conference (MobiMedia 2009), EUMOB Workshop, London, UK, September 2009.

[38] [DNAICSSA08]
N. Mbarek, F. Krief, D. Négru, "SLNP Usage for QoS Negotiation in Heterogeneous Environments", in proc. of the 6th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA 2008), WiSe Workshop, Doha, Qatar, March 2008.

[39] [DNGLOB06]
D. Négru, T. Ahmed, A. Mehaoua, "Ambient Middleware Framework for Efficient Service Provision to Users Anywhere around the Globe", in proc. of the 2006 IEEE Global Communications Conference (GLOBECOM'06), San Francisco, California, USA, December 2006.

[40] [DNWCNC06]
D. Négru, T. Ahmed, A. Mehaoua, "Adaptive audio-video streaming solution over IP mobile environments", in proc. of the IEEE Wireless Communications and Networking Conference (WCNC 2006), Las Vegas, Nevada, USA, April 2006.

[41] [DNWCNCI06]
I. Djama, T. Ahmed, D. Négru, N. Zangar, "An MPEG-21-enabled Video Adaptation Engine for Universal IPTV Access", in proc. of the IEEE Wireless Communications and Networking Conference (WCNC 2006), ISBC Symposium, Las Vegas, Nevada, USA, April 2006.

[42] [DNTRID06]
D. Négru, A. Mehaoua, E. Pallis, "ATHENA: a Large-Scale Testbed for the Next Generation of Interoperable Networks and Services", in proc. of the 2nd International IEEE/Create-Net Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities (TridentCom 2006), Barcelona, Spain, March 2006.

[43] [DNICC06]
Y. Hadjadj-Aoul, D. Négru, A. Mehaoua, "Predictive Resource Management for mixed IP/DTV services in Broadcasting Networks", in proc. of the International Conference on Communications (ICC 2006), Istanbul, Turkey, June 2006.

[44] [DNIST05]
D. Négru, A. Mehaoua, E. Pallis, “MIPv6 Extensions for seamless IP Mobility in Converged DVB-T/WLAN Networks”, in proc. of the 2005 IST Mobile & Wireless Communications Summit, Dresden Germany, June 2005.

[45] [DNGLOB04]
Y. Hadjadj-Aoul, A. Nafaa, D. Négru, A. Mehaoua, “FAFC: Fast Adaptive Fuzzy AQM Controller for TCP/IP Networks”, in proc. of the 2004 IEEE Global Communications Conference (GLOBECOM’04), Dallas, Texas, December 2004.

Tutoriaux dans conférences internationales (avec comité de lecture):

[46] Eugen Borcoci, Daniel Negru, “Content Oriented Routing and Forwarding”, in NexComm 2012, Chamonix, France, Mai 2012.

[47] Eugen Borcoci, Daniel Negru, Christian Timmerer, “Future Internet Trends: Networked Media, Content and Services Orientation”, in INFOWARE 2011, Luxembourg, Luxembourg, June 2011.

[48] Eugen Borcoci, Daniel Negru, Christian Timmerer, “Content Aware Networking: a Future Internet Perspective”, in NetWare 2010, Venice, Italy, July 2010.

Contributions de standardisation:

[49] ETSI MCD CDNI Working Group: “ETSI TS 102 990: CDN Interconnection, use cases and requirements (Use_case_and_requirement_from_FP7_Alicante)”, Accepted as part of the official standard document. Joint contribution in collaboration with Orange FT, University College London, Synelixis, March 2012.

[50] IETF CDN-Interconnection (CDNI) Working Group: “Catalogue of Advanced Use Cases for Content Delivery Network Interconnection”, Accepted as draft: draft-fmn-cdni-advanced-use-cases. Joint contribution in collaboration with EC FMN cluster, October 2011.

[51] Broadband Forum (BBF) Marketing Group & End-to-end architectures Working Group: “ALICANTE: Network, content and context awareness in media delivery”, Presentation performed. Accepted for official White Paper elaboration, May 2011.

[52] Future Internet Assembly (FIA) group: “Media Ecosystems: A Novel Approach for Content-Awareness in Future Networks”, Accepted as part of the official FIA recommendations in ‘Content’ section, March 2011.

[53] MPEG Advanced IPTV Terminal (AIT) group: “Response to the AIT 2nd CfP: Authenticate User and Identify User Elementary Service”, Accepted for presentation at MPEG meeting., August 2010.

2. Encadrement doctoral et scientifique

Tableau Récapitulatif

Type	Thèses	Post-Doc	Ingénieur de Recherche	Stagiaires
Nombre	6 (5 soutenues)	2	1	8
Dates	1 : 2008-2011 2 : 2008-2012 3 : 2010-2013	1 : 2010-2013 1 : 2011-2012	2009-2013	2008-2011
Nb publications	23	5	2	-

<u>Encadrement de 6 thèses :</u>

1. **AIT-CHELLOUCHE Soraya** : Doctorante de l'Université de Bordeaux 1, laboratoire LaBRI, 2008-2011, financement sur bourse MENRT (BQR).

Co-encadrement : Daniel Négru (90%), Pr. Francine Krief (10%).

Thèse soutenue le 09 Décembre 2012, mention Très Honorable.

Sujet de thèse : Délivrance de services médias suivant le contexte au sein d'environnements hétérogènes pour les réseaux médias du futur.

Publications émanant du travail de thèse : 5 dans conférences internationales avec comité de lecture.

2. **ARNAUD Julien** : Doctorant de l'Université de Bordeaux 1, laboratoire LaBRI, 2008-2012, financement sur contrat de recherche industriel.

Co-encadrement : Daniel Négru (90%), Pr. Francine Krief (10%).

Thèse soutenue le 17 Décembre 2012, mention Très Honorable.

Sujet de thèse : Gestion de mobilité efficace et contexte-dépendante dans les réseaux télécoms de nouvelles générations.

Publications émanant du travail de thèse : 5 dans conférences internationales avec comité de lecture, 1 dans journal avec comité de lecture.

3. **AUBRY Willy** : Doctorant de l'Université de Bordeaux 1, laboratoires LaBRI et IMS, 2008-2012, thèse en co-direction et pluri-disciplinaire Informatique et Electronique, financement sur bourse CIFRE et projet ANR.

Co-encadrement : Daniel Négru (40%), Dr. Bertrand Le Gal (40%), Pr. Dominique Dallet (10%), Pr. Francine Krief (10%).

Thèse soutenue le 19 Décembre 2012, mention très Honorable.

Sujet de thèse : Conception en co-design logiciel/matériel d'une plateforme reconfigurable d'adaptation vidéo pour les réseaux du futur.

Publications émanant du travail de thèse : 4 dans conférences internationales avec comité de lecture, 1 dans journal avec comité de lecture.

4. **RODRIGUES Preston** : Doctorant de l'Université de Bordeaux 1, laboratoire LaBRI, 2010-2013, thèse pluri-disciplinaire Réseaux et Langages, financement sur projet européen.

Co-encadrement : Daniel Négru (30%), Dr. Laurent Réveillère (30%), Dr. David Bromberg (30%), Pr. Francine Krief (10%).

Thèse soutenue le 27 Mai 2013, mention Honorable.

Sujet de thèse : Approche langage pour l'adaptation dynamique des flux multimédia au sein de l'Internet du Futur.

Publications émanant du travail de thèse: 3 dans conférences internationales avec comité de lecture.

5. **CHERIF Wael** : Doctorant de l'Université de Rennes 1, laboratoires IRISA-INRIA et LaBRI, 2010-2013, thèse en co-direction, financement sur bourse CIFRE.

Co-encadrement : Daniel Négru (30%), Dr. Adlen Ksentini (60%), Pr. Gerardo Rubino (10%).

Thèse soutenue le 19 Juin 2013, mention très Honorable.

Sujet de thèse : Adaptation de contexte basée sur la Qualité d'Expérience dans les réseaux Internet du Futur.

Publications émanant du travail de thèse: 3 dans conférences internationales avec comité de lecture, 1 dans un magazine international avec comité de lecture.

6. **ANAPLIOTIS Petros** : Doctorant de l'Université de Bordeaux 1 et de Technological Educational Institute of Crete (TEIC), laboratoires LaBRI et PASIPHAE, 2010-2013, thèse en co-direction, financement sur projet européen.

Co-encadrement : Daniel Négru (30%), Dr Evangelos Pallis (30%), Dr Georgios Gardikis (30%), Pr. Francine Krief (10%).

Soutenance prévue : Octobre 2014.

Sujet de thèse : Routage basé sur le contenu dans les réseaux média du futur.

Publications émanant du travail de thèse (en cours) : 3 dans conférences internationales avec comité de lecture.

Encadrement de 2 chercheurs post-doctorants :

1. **CHEN Yiping** : Docteur de l'Université de Rennes 1.

Objectif : réalisation de travaux sur l'évolution architecturale des réseaux médias du futur, basés sur le contenu.

Encadrement : 100%

Durée : 3,5 ans (2010-2013).

Financement sur projet européen.

Publications émanant du travail de Post-Doc (en cours) : 4 dans conférences internationales avec comité de lecture, 1 dans journal avec comité de sélection.

2. **AIT CHELLOUCHE Soraya** : Docteur de l'Université de Bordeaux 1.

Objectif : réalisation de travaux sur des solutions de gestion du contexte utilisateur basées sur les ontologies et de distribution efficace de services médias en fonction de ce contexte.

Co-encadrement : Daniel Négru (70%), Dr. Adlen Ksentini (30%)

Durée : 1 an (2011-2012).

Financement sur projet européen.

Travail en collaboration avec le laboratoire IRISA (équipe INRIA Dionysos) de l'Université Rennes 1.

Publications émanant du travail de Post-Doc (en cours) : 2 dans journaux avec comité de sélection, 1 dans conférence internationale avec comité de lecture.

Encadrement d'un Ingénieur de Recherche :

1. **DESFARGES Simon** : Ingénieur de l'Institut Polytechnique de Bordeaux (ENSEiRB-MATMECA).

Objectif : développement de modules logicielles/matériels pour adaptation vidéo dans un contexte de Home Networking (2009-2013).

Co-encadrement : Daniel Négru (50%), Dr. Bertrand Le Gal (50%)

Durée : 3,5 ans (2009-2013).

Financement sur projet ANR.

Travail en collaboration avec le laboratoire d'électronique IMS (équipe COFI) de l'Université de Bordeaux 1.

+ Encadrement de 8 stagiaires (M2 recherche et école d'ingénieur)

3. Responsabilités

Tableau Récapitulatif

	Projets	Collaborations directes	Comités	Administration & Enseignement
Nombre	1 projet collaboratif IP FP7 en cours; 1 ANR ARPEGE terminé ; 6 projets collaboratifs terminés ; 2 projets industriels (Thomson, Sogeti). + 1 projet CHIST-ERA à démarrer (01/2014) + 1 projet collaboratif IP FP7 à démarrer (12/2013)	10 collaborations en cours (7 avec labos de recherche France/Europe, 3 avec industriels).	10 organisations de conférences (workshops, sessions spéciales ou édition de journaux).	4 responsabilités en cours à l'ENSEIRB-MATMECA (dpt Télécom).
Faits significatifs	Coordination générale + scientifique et technique du projet européen FP7 ALICANTE (20 partenaires).	+ de 30 publications émanant de ces collaborations.	Au sein de conférences reconnues : ICME, ICECS, EUSIPCO…	En charge de responsabilités depuis 2006.

Responsabilités scientifiques – Projets & Collaborations:

I- Projets

1. **Projet européen : ALICANTE** (Media Ecosystem Deployment through Ubiquitous Content-Aware Network Environments – http://www.ict-alicante.eu) :
   - Projet Intégré (IP) du 7e programme-cadre de la commission européenne dans le domaine des Technologies de l'Information et de la Communication (ICT FP7) ;
   - Projet collaboratif autour de 20 partenaires (industriels, PME, académiques) ;
   - Budget : 10,5 Millions € ;
   - Dates : 2010-2013 (42 mois) ;
   - Objectif : proposer une solution innovante, adéquate et standardisée répondant aux problèmes de l'évolution des réseaux de média au sein de l'Internet du Futur, vers une approche consciente du contenu et du contexte (Content- and Context-Aware Network) et des applications conscientes des caractéristiques réseaux (Network-Aware Applications) ;
   - Responsabilités : Coordinateur du projet, Responsable scientifique et technique du projet.

2. **Projet ANR : ARDMAHN** (Architecture Reconfigurable Dynamiquement et Méthodologie pour l'Auto-adaptation Home Networking – http://ardmahn.org) :
   - Projet ANR du programme ARPEGE – systèmes embarqués ;
   - Projet collaboratif autour de 5 partenaires (industriels, PME, académiques) ;
   - Budget : 2,9 Million € ;
   - Dates : 2009-2013 (42 mois) ;

- Objectif : proposer une solution innovante pour pallier aux problèmes d'hétérogénéité des terminaux, services et réseaux sur toute la chaine multimédia. L'adaptation dynamique et transparente des flux multimédias selon les caractéristiques clientes est la clé du succès des services réseaux de demain. Pour y arriver, il est nécessaire d'étudier sa mise en place au niveau des réseaux en lien direct avec les utilisateurs (le home network) ;
- Responsabilités : Responsable scientifique au niveau du partenaire LaBRI-IMS.

3. **Projet industriel Sogeti-LaBRI : Multibox** (Réalisation de services composés multimédias en réseaux et mise en place d'une plateforme de démonstration sous systèmes hétérogènes):
   - Projet dans le cadre du partenariat Sogeti-LaBRI ;
   - Dates : 2011-2012 (18 mois) ;
   - Objectif : mettre en œuvre une solution client-serveur pour la distribution efficace de flux multimédias, permettant la composition et l'adaptation de services (IPTV+données, Vidéo-conférence+données, etc…) vers clients hétérogènes au sein de la maison et en dehors, tout en tenant en compte le contexte utilisateur et les contraintes des ressources des équipements ;
   - Responsabilités : Responsable du projet au niveau du LaBRI.

4. **Projet industriel Thomson-LaBRI : ConvergenceIMS** (Convergence des services pour les environnements hétérogènes dans les réseaux de Nouvelle Génération IMS):
   - Contrat de prestation de recherche avec Thomson Grass Valley, Rennes ;
   - Budget : 148 K€ ;
   - Dates : 2008-2011 (36 mois), terminé ;
   - Objectif : étude et mise en place d'une nouvelle solution de diffusion vidéo adaptative au sein d'un réseau IMS avec clients hétérogènes, basée sur de nouveaux algorithmes de décision ;
   - Responsabilités : Responsable du projet au niveau du LaBRI.

+ Participation à plusieurs projets collaboratifs européens et ANR :
- FP7 ICT ADAMANTIUM (2008-2011) ;
- FP7 ICT SEA (2008-2010) ;
- FP6 IST ENTHRONE 1 & 2 (2004-2008) ;
- FP6 IST IMOSAN (2006-2008) ;
- FP6 IST ATHENA (2004-2006) ;
- ANR RIAM NMS (2003).

+ Réponses à de nombreux appels à projets :
- FP7 ICT : Call 1-8 Strategic Objectives 1.1, 1.5, 3.7 ;
- ANR : ARPEGE, VERSO ;
- EUREKA CELTIC, EUROSTARS.

+ 1 projet CHIST-ERA à démarrer 01/2014 sur les aspects adaptation suivant le contenu;
+ 1 projet FP7 IP T-NOVA à démarrer 11/2013, sur les aspects virtualisation et cloud pour le media ;

II- Collaborations directes

1. **Collaboration avec le laboratoire IMS, Bordeaux (équipe COFI),** sur un projet autour du co-design matériel/logiciel pour l'adaptation multimédia:
   - Co-encadrement d'une thèse (W. Aubry) ;
   - Co-encadrement d'un ingénieur de recherche (S. Desfarges) ;
   - Participation commune à un projet ANR (ARDMAHN) ;
   - Publications émanant de la collaboration : 5 (+1 en cours).

2. **Collaboration avec le laboratoire IRISA, Rennes (équipe INRIA-Dionysos),** sur un projet autour de la QoE, outils, méthodologies et prototypage (A-PSQA QoE Tool):
    - Co-encadrement d'une thèse (W. Chérif) ;
    - Co-encadrement d'un post-doc (S. Ait Chellouche) ;
    - Publications émanant de la collaboration : 2 (+2 en cours).

3. **Collaboration avec le laboratoire PASIPHAE, Grèce,** sur plusieurs projets autour de la convergence des réseaux et services à tous niveaux :
    - Co-encadrement d'une thèse (P. Anapliotis) ;
    - Validation en cours du financement d'un projet (VELLEROFONTIS) ;
    - Participation commune à plusieurs projets européens (ATHENA, ALICANTE) ;
    - Montage de conférences, workshops et écoles d'été ;
    - Publications émanant de la collaboration : 8.

4. **Collaboration avec le laboratoire DEMOKRITOS, Grèce (équipe Medianet),** sur plusieurs projets autour des réseaux de média et conscients du contenu :
    - Participation commune à plusieurs projets européens (ATHENA, IMOSAN, ADAMANTIUM, ENTHRONE, ALICANTE) ;
    - Montage de conférences, workshops ;
    - Editions de journaux ;
    - Publications émanant de la collaboration : 7.

5. **Collaboration avec l'Université de Klagenfurt, Autriche (équipe MMC-ITEC),** sur plusieurs projets autour de la diffusion et l'adaptation multimédia (DASH-VLC encoder-decoder-streamer, DASH proxy gateway, SVC Encoder):
    - Participation commune à plusieurs projets européens (ENTHRONE, ALICANTE) ;
    - Montage de conférences, workshops ;
    - Editions de journaux ;
    - Publications émanant de la collaboration : 6.

6. **Collaboration avec l'Université Polytechnique de Bucarest, Roumanie (équipe ET),** sur plusieurs projets autour de la virtualisation des réseaux, basée sur le contenu :
    - Participation commune à plusieurs projets européens (ATHENA, ENTHRONE, ALICANTE) ;
    - Montage de conférences, workshops, sessions spéciales et tutoriaux ;
    - Publications émanant de la collaboration : 10.

7. **Collaboration avec le laboratoire LEAT de l'Université de Nice,** sur un projet autour de la reconfiguration dynamique partielle au sein de FPGA :
    - Participation commune à un projet ANR (ARDMAHN) ;
    - Publications émanant de la collaboration : 1.

8. **Collaboration avec l'entreprise Thomson Grass Valley/Thomson Video Networks, Rennes,** sur plusieurs projets autour des technologies de routage en fonction du contenu, d'adaptation vidéo et des réseaux télécom de nouvelles générations, prototypage (MANE – CAN Router, Convergence IMS adaptive gateway) :
    - Participation commune à des projets européens et ANR ;
    - Montage de projets industriels directs ;
    - Réalisation de prototypes logiciels et matériels ;
    - Montage de conférences, workshops ;
    - Publications émanant de la collaboration : 6.

9. **Collaboration avec la PME Viotech Communications, Paris,** sur plusieurs projets autour des Home Gateway et leurs extensions, prototypage (Home-Box) :
   - Participation commune à des projets européens et ANR ;
   - 2 thèses sur financement CIFRE (W. Aubry, W. Chérif) ;
   - Réalisation de prototypes logiciels et matériels ;
   - Publications émanant de la collaboration : 11.

10. **Collaboration avec l'entreprise Sogeti, Pessac (Cellule Innovation)** : voir projet ci-dessus.

Responsabilités scientifiques – Comités:

Organisation de conférences, workshops, sessions spéciales, édition de journaux :

- Conférence TEMU 2010 et 2012 et 2014 (www.temu.gr);
- Hindawi Journal of Computer Networks and Communications, special issue on "Advanced Techniques for Multimedia Delivery in Collaborative Networks", 2012 (http://www.hindawi.com/journals/jcnc/si/830720/cfp/);
- Special Session MDCAN@EUSIPCO 2012 (http://www.eusipco2012.org/program_special_sessions.php);
- Workshop WoMAN@ICME 2011 (http://woman2011.ict-alicante.eu/);
- Special Session VOTN@ICECS 2011 (http://icecs2011.com/);
- Hindawi International Journal of Digital Multimedia Broadcasting (IJDMB), special issue on "IP and Broadcasting Systems Convergence", 2010 (http://www.hindawi.com/journals/ijdmb/si/832141/cfp/);
- Special Session AMD@TEMU 2010 (www.temu.gr);
- Special session "Networks & Applications" @ AQTR 2008;
- Workshop WISE@AICSSA 2008.

Membre du comité de programme des conférences et journaux suivants:

- WCNC, ICC, EuroITV, TEMU, Globecom, EUSIPCO, VTS, IARIA, MobiMedia;
- Springer's Wireless Personal Communications Journal, Hindawi Journals.

Membre élu du Comité Recherche de l'ENSEIRB-MATMECA depuis 2010.

Membre du Comité de Sélection de Recrutement de Maitre de Conférences :

- Université de Versailles St Quentin en Yvelines, poste 27 MCF 4043, 2012 ;
- Institut Polytechnique de Bordeaux (IPB), poste 27 MCF 1345, 2010.

Membre du jury de thèse de N. Nouri, Institut National Polytechnique de Lorraine, Nancy-Université, « Evaluation de la qualité et transmission en temps réel de vidéos médicales compressées : application à la télé-chirurgie robotisée », Septembre 2011.

Responsabilités administratives et d'enseignement:

Responsable de la matière « Réseaux » au sein du département Télécom de l'ENSEIRB-MATMECA, IPB.
Gestion de l'ensemble du cursus « Réseaux » sur les 3 années d'enseignement.

Responsable option 3^e^ année Telecom à l'ENSEIRB-MATMECA, IPB.
Nom de l'option : Réseaux et Systèmes embarqués Communicants (RSC).

Responsable des « Stages » de 3e année (PFE) des étudiants au sein du département Telecom à l'ENSEIRB-MATMECA, IPB.
Soutien dans la recherche des stages, gestion des stagiaires, évaluations, …

Mise en place d'une école d'été sur le thème des « Réseaux pervasifs » 2012 et 2013 (http://www.pasiphae.eu/perseus/)
Financement sur un programme européen.
5 étudiants du département sélectionnés.

Charges d'enseignement :

- 2007-2008 : 304 HeqTD;
- 2008-2009 : 268 HeqTD;
- 2009-2010 : 233 HeqTD;
- 2010-2011 : 203 HeqTD;
- 2011-2012 : 208 HeqTD;
- 2012-2013 : 301.5 HeqTD.