Available online at www.sciencedirect.com
ScienceDirect
Computer Communications 30 (2007) 638–651
computer communications
www.elsevier.com/locate/comcom

End-to-end quality of service provisioning through an integrated management system for multimedia content delivery

Toufik Ahmed [a,*], Abolghasem (Hamid) Asgari [b], Ahmed Mehaoua [c], Eugen Borcoci [d], Laure Berti-Équille [e], Kormentzas Georgios [f]

[a] CNRS-LaBRI Laboratory, University of Bordeaux-1, Bordeaux, France
[b] Thales Research and Technology (TRT) Limited, Worton Drive, Reading RG2 0SB, UK
[c] CNRS-PRiSM Laboratory, University of Versailles, Versailles, France
[d] University "Politehnica" of Bucharest, Bucharest, Romania
[e] INRIA – IRISA, Rennes, France
[f] Institute of Informatics and Telecommunications, Demokritos, Athens, Greece

Available online 20 November 2006

Abstract

Next generation networks will be complex, interconnecting different technologies, and architectures (IP, DVB-T/S, UMTS, GSM/GPRS, etc.) across a multitude of software and hardware platforms for offering a large number of value-added services including multimedia, audio–visual, etc. One of the major requirements for the successful and wide deployment of such services is the efficient transmission of performance sensitive digital items (i.e., audio, video, and image) over a broad range of bandwidth-constrained wired/wireless access/core/distribution networks. This paper presents an approach to the management and provisioning of end-to-end Quality of Service (QoS) in next generation networks (IP and non-IP networks) that is being developed in the IST ENTHRONE[1] project. The primary goal of this project is to provide a solution for seamless access to multimedia content with end-to-end QoS support through integrated management of content, networks and terminals. The adopted methodology in providing end-to-end QoS is through an integrated management in the context of MPEG-21 framework for managing services at both customer and provider levels using Service Level Agreements (SLA) concept and provisioning network resources for satisfying the customer QoS demands as well as meeting the QoS requirements of multimedia contents.

Keywords: Integrated Management System (IMS); Middleware; End-to-end QoS; MPEG-21; IP; Diffserv; DVB

1. Introduction

The Internet, is moving from being a simple monolithic data service network to a ubiquitous multi-service network in which different stakeholders including but not limited to content providers (**CPs**), service providers (**SPs**), and network providers (**NPs**) require to co-operate for offering value-added services and applications to content consumers (**CCs**). These services and applications may dynamically be created and managed.

The problem of how to extend QoS capabilities across multiple provider domains for providing end-to-end services has not been solved satisfactorily to-date. Furthermore, developing a unified solution that enables the end-to-end delivery of services over various types of networks at a guaranteed quality level is more challenging. This problem is augmented by the presence of various business actors (CP, SP, NP) and the customers' new service requirements. The current practice in service offering is through the use of Service Level Agreements (**SLAs**). The technical part of the

* Corresponding author.
E-mail address: tad@labri.fr (T. Ahmed).

[1] IST FP6 ENTHRONE project is partially funded by the Commission of the European Union. For more information visit http://www.enthrone.org/.

0140-3664/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2006.10.009

SLA is the SLS (Service Level Specification). TEQUILA defined a set of semantics for SLSs as the basic component for defining QoS-based IP connectivity services [1]. SLAs describe the characteristics of the service offering and the responsibilities of the parties involved for using/providing the offered service. SLAs between consumer and services providers (**cSLAs**) are used to formulate the customer requests as well as customers' and providers' service commitments. Provider level service agreements (**pSLAs**) are between providers, i.e., SPs and NPs and between NPs themselves. These agreements are supposed to lay the first stone for providers towards extending their geographical scope beyond their boundaries. This is the first challenge in offering inter-domain QoS delivery.

Developing a common framework to bridge the gap between stakeholders would facilitate cooperation and support a more efficient implementation and integration of the different needs and requirement for each one. In this respect, **MPEG-21** aims to create the big picture of a common framework to guarantee interoperability by focusing on how the elements of a multimedia application infrastructure should relate, integrate, and interact. The MPEG-21 data model will be used to provide the common support for implementing the required functionality and managing the resources.

The provision of an Integrated Management System (IMS) middleware for end-to-end QoS delivery across heterogeneous networks and terminals can be considered to be a key element in the successful mass market deployment of audio–visual services that would produce revenues for the content/service providers as well as network providers.

This paper presents the solution currently being developed and implemented in the ENTHRONE project for end-to-end QoS-enabled service delivery over heterogeneous networks. The IMS architecture is based on a distributed middleware model, where diversified QoS policy based management functions are performed in many geographically distributed environments (content generation, networks, and terminals). A major goal of this project is to bridge the divide between the content provision and the networking worlds, resulting in cross-industry co-ordination on both network and content management issues, and bringing the focus to mutually advantageous standards such as MPEG-21.

This paper is organized as follows. Section 2 highlights some related works and briefly describes the ENTHRONE focus. Section 3 explains how the service requests are formulated and highlight the role of IMS middleware for end-to-end QoS offering. Section 4 describes how multimedia content management is achieved. Section 5 presents the approach for end-to-end QoS provisioning over heterogeneous networks. Section 6 presents the overall functional architecture and the required protocols to be used in the multi-domain networking environment. Finally, conclusions are presented in Section 7.

2. Related work

In contrast to the significant research work carried out for providing intra-domain QoS solutions, limited work has been conducted on providing inter-domain QoS solutions. Many researchers have considered inter-domain traffic engineering for best effort traffic [2]. Using SLA policies to enable IP network providers agree how to distribute service-level constraints (i.e., QoS) across multiple domains is proposed in [3]. However, it should be noted that majority of these contributions are solutions without an integrated and complete view of how to deliver QoS across multiple heterogeneous domains at large and in a scalable way. There have also been a number of IST European research projects [4]), which have addressed mainly QoS-based connectivity services across IP networks.

AQUILA implemented a QoS-based architecture for controlling, monitoring, and accessing to the resources in DiffServ networks [4]. In view of service offering, AQUILA did not take into account any business-related or technical layer aspects in service offering such as SLAs, business processes, etc. It did not consider the roles of different stakeholder roles at service, access, and network levels. It only uses a service creation process (definition of end-user services by providers) using a tool kit.

The CADENUS developed an architecture, which includes functional blocks at the user-provider interface, within the service provider domain, and between the service provider and the network provider [4,5]. It defined three key components: Access Mediator (AMd), Service Mediator (SMd), and Resource Mediator (RMd). The CADENUS architecture does not get into details of how static and dynamic resource management and traffic engineering is achieved at the network level. CADENUS system allows auditing of the SLA with service provider and the dynamic re-negotiation of the service agreements.

The CADENUS solution involves a significant amount of signalling to process an end customer request [6]. An initial dialogue is performed between the end user and AMd in order to select a service from a list offered by AMd. After this selection is done, a new set of parallel dialogues have to be done between AMd and one or several candidate SMd(s). Each SMd in turn, has to contact the first RM in the chain, (and the latter may have to contact others in the inter-domain chain) based on SLSs, to allow RMs(s) to make an evaluation of the impact the service is going to have on the network and derives a cost to be paid for service enforcement. This operation is performed by means of an admission control process (RMd function), which implies the participation of all the network management entities. Note that some resources are committed (pre-allocated) in each RMd which replied "yes". The computed cost is returned to each SMd by its corresponding RM. The SMd(s) return the results to AMd, which sorts them and offers the list to the end user. After end user response/selection only one SMd is eventually chosen and a transaction rollback should be performed by the AMd for those SMd(s) which have been rejected. The ENTHRONE design philosophy, as explained in the following sections, avoids such a complicated set of parallel transactions AMd – SMd_k, $k \geqslant 1$ and also avoids pre-allocation/rollback actions in different RMd(s).

The objective of the TEQUILA project was to study, specify, implement, and validate a set of service definition and traffic engineering tools to obtain quantitative end-to-end QoS guarantees [4]. The QoS architecture in TEQUILA is decomposed [7] at a high level, into the following major subsystems: Service Management [8] Traffic Engineering [7] and Monitoring subsystems [9]. TEQUILA focused on operational management of IP connectivity services across a single domain both at the service and resource levels.

MESCAL is TEQUILA's follow-up project focusing on the support of QoS across multiple IP domains [4] at both service and resource levels. The QoS functional architecture in MESCAL is decomposed into the following functional groups: Service Definition and QoS Capabilities Exchange, Traffic Engineering, SLS Management, Monitoring and Assurance, and Traffic Enforcement [10]. At the service level, neither TEQUILA nor MESCAL is concerned with end-user services or service creation, i.e., the process of (automated) service definition and service offering by the Service Providers. They are mainly focused on QoS enabled IP connectivity services. Neither project developed an end-user application service framework taking into account the business-related aspects of high-level service offerings and roles of content and service providers. This is an area where ENTHRONE is positioned to contribute on.

The objective of the ongoing EuQoS IST project [11] is to integrate and validate end-to-end QoS networking technologies to support advanced QoS-aware multimedia applications over multiple, heterogeneous network domains. The focus is on the development of an integrated resource management system called EuQoS System and an associated EuQoS signalling protocol. The EuQoS limits its scope only to end-to-end network resource management.

ENTHRONE differs from the stated IST projects in that its end to end scope and business model encompasses all relevant actors including content providers, service providers, network providers, and content consumers. However, ENTHRONE builds on the results of these IST projects where appropriate.

2.1. Enthrone focus

ENTHRONE covers an entire audio–visual service distribution chain, including content generation and protection, distribution across QoS-enabled heterogeneous networks, and delivery of content at user terminals. In ENTHRONE, the whole content delivery chain is considered, including three distinct providers: the content provider, the network provider, and the service provider.

Prior to fulfilling the customer service requests, resource provisioning is performed in advance. To this aim, the IMS middleware at SP negotiates and concludes two types of contracts: pSLAs with CPs in order to enable the SP to offer contents to the customers; pSLSs with NPs in order to establish aggregated QoS enabled pipes through inter-domain from Contents Servers region towards regions where potential Content Consumers are located (the necessity of such pipes comes out from forecast data on users and required services. This approach creates a scalable solution because it avoids to go through the chain of Service Provider to Network Provider and between Network Providers to provision the involved networks, for every requested service. On the other hand the ENTHRONE solution of in-advance-resource-provisioning is not a static. The IMS architecture includes a feedback loop including the Sevice Management and Resource Management (inter and intra-domain). A resource provisioning cycle involving network resources re-dimensioning is possibly to be run at medium-term intervals (the frequency of such cycles is established by each NP management). Another dynamic feature consists of the possibility to activate the pSLS-based aggregated pipes in a dynamic manner depending on the current needs (using the EQoS-RA protocol, see the Section 6.5 for details).

The MPEG-21 data model is used to provide the common support for implementing and managing the resources' functionality. ENTHRONE is concerned with end-to-end QoS in terms of performance, at both the user and network levels, and with mapping between these levels.

ENTHRONE design philosophy makes separation between service and transport/resource management. This is an important architectural choice allowing a logical integration of service management irrespective of underlying transport mechanisms in mono- or multi-domain environments. Also the networking infrastructures and technology are not directly visible at the service level, therefore, the heterogeneous infrastructures are naturally allowed by the architecture. The ENTHRONE adopted solutions thus follows the recent trend recognized in Next Generation Networks (NGN) architectures – recently proposed as a "merge" between IP and telecom worlds.

With regard to managing the services and the network resources and their related areas, the technical focus of the ENTHRONE project is highlighted below:

- To provide content generation and protection mechanisms.
- To create QoS-enabled services for content distribution. Toward this end, we develop a customer/end-user service framework, defining, and offering services through Service Providers.
- To define and implement of MPEG-21 user terminals, where the service requests are initiated.

- To manage the whole chain of content delivery from Content Servers to the end-user terminals. To this aim, we manage the service level-related co-operation and dialogue between different stakeholders (CP, SP, NP, CC) in the content delivery chain.
- To provide service level monitoring (based on monitoring agents at node levels and network monitors at both intra- and inter-domain as well as user profile and device characteristics at terminals). Additionally, to develop and utilise Quality Meters able to measure the perceived quality level of an audio–visual stream as part of service level monitoring.
- To manage end-to-end QoS delivery chain including both IP and non-IP networks in network provisioning and configurations and the user terminal reception, using MPEG-21 features.

3. Integrated Management System and its role

We propose an Integrated Management System (**IMS**) as a deployed middleware for service management (provisioning, service requests handling, service assurance, multimedia content management using descriptive metadata), and multi-domain resource management. The service requests and request handling is accomplished using service level agreement and specification introduced for DiffServ-enabled networks. IMS is service oriented architecture, following a distributed approach. A specific IMS middleware (IMS subsystem) is proposed for every actor on the service delivery chain in view of fulfilling the requirements of each entity/organization involved in service offering. As a consequence, the IMS has a number of functional components for each entity/organization. However, the IMS functional components listed below can be present or absent in each organization/entity, depending on the role of entity (i.e., SP, CP, CC, and NP) in the end-to-end service delivery chain. The high-level IMS architecture is shown in Fig. 1 and its subsystems are briefly described below:

IMS Dispatcher and especially its **Service Manager** positioned at the service provider entity is responsible for handling customer subscriptions (cSLAs), conducting contracts with CPs (pSLAs) and NPs (pSLSs), provisioning and offering services, and granting the access to the services, which has been chosen by the customers. These are limited to the services owned by the SP. In addition, it includes a **Metadata Manager** describing and handling the multimedia content. IMS Dispatcher also performs Digital Item Adaptation processing. An Adaptation Decision Engine (ADE) takes adaptation decisions with the goal of providing customers a service with the best quality given constraints on the terminal, the network and the cost of service.

Content Manager is responsible to keep information related to the content, and to support all operations related to any subscriptions for the use of content, the management of and interaction with the content.

Terminal Device Manager (**TDM**) is responsible for communication with the terminal device by providing appropriate IMS interface to various kinds of terminals. TDM also retrieves the terminal capabilities and the characteristics of the access network to which the terminal is connected for delivering this information to the IMS Dispatcher. In addition, TDM is involved in a number of tasks including QoS level Monitoring and License Handling.

Network Manager (**NM**) contains a facility for handling inter-domain issues in terms of QoS discovery, QoS-based route selection and the different types of QoS services that can be offered. The Network Manager has some knowledge of the resources and control layers, with the inter-domain

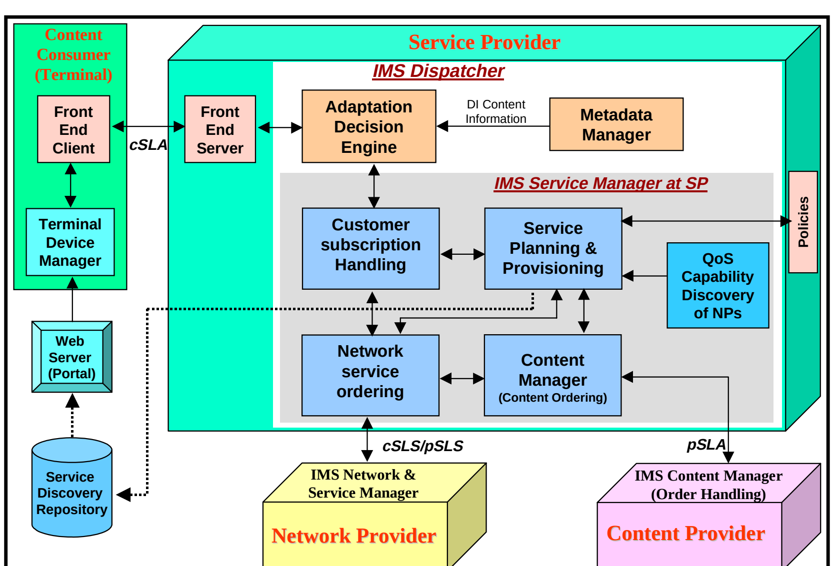

Fig. 1. IMS architecture.

vision for QoS service set-up in advance. There is an IMS compatible Service Manager at each NP, which handles SP's as well as other NP's service (i.e., pSLS) requests [12]. It also orders any necessary pSLSs to adjacent NPs.

Each domain has an intra-domain **Resource Manager** (**RM**) which controls and manages the resources inside the domain to support QoS.

Initially the SP establishes pSLA contract with one or several CPs for getting the access rights to Digital Items and for offering these contents to its customers. Then, the SP is informed by a CP about the location(s) of its Content Servers (CS) and some details about the contents. The SP also foresees through its Service Planning and provisioning functional block, the locations (regions) where groups of its potential customers are placed. In this approach the SP constructs aggregated pipes in co-operation with the NPs involved in the data path between Content Servers and Content Customer regions. A virtual network of such aggregated pipes (traffic trunks) is established at the inter-domain level. Therefore, prior to fulfilling the customer service requests, resource provisioning is performed in aggregated level in advance. This creates a scalable solution as (1) the operational cost is reduced since there is no need to go through the whole chain of SP, CP, NPs, to provision the involved networks, for every customer request. (2) A large population size can be serviced in a timely fashion.

After the associated pipes (being the means for distributing contents from CPs to several destinations) are put in place, the SP can offer services to customers/users.

An end-user selects a service with a given or chosen quality, by means of either negotiating SLAs dynamically or accepting generic SLAs offered by SP. The motivation for using SLAs and SLSs is to provide assured end-to-end quality of service to the end-user in a simple and efficient manner.

It is assumed that a SP offers its services via a "Front-End" facility. In addition, it may advertise its services through a portal where a customer's request is redirected to the appropriate SP Front-End facility. The portal allows the user to select a service with a given or chosen quality of service. The IMS Dispatcher receives a cSLA request initiated by the end user (i.e., video on demand with gold service) and relayed by the front-end. After performing some initial checks and adaptation, it passes the adapted cSLA request to the Service Manager module of the IMS Dispatcher. The Service Manager translates the customer cSLA into SLSs taking into account the related metadata information and any adaptation decision made. After retrieving the service availability from an SLS repository, the Service Manager performs admission control to decide whether to accept or reject the request. SLS enforcement is achieved by means of a policy-based approach as explained in the next sections.

All IMS subsystems are viewed as a collection of middleware components. The communication between the middleware components can be achieved by using standard Distributed Objects Technologies (DOT) such as (RMI/CORBA-IIOP/SOAP) where appropriate, or more specialized communication protocols can be used to offer a better efficiency. As an example, an IMS distributed implementation could use DOT for communication between some subsystems like SM, NM/RM, and Monitoring System if they are remote located from each other. Middleware communication protocols can also be used horizontally at a specific layer or vertically for cross-layer interactions. Specialized standard communication protocols such as Common Open Policy Service (COPS) are better fitted to vertical communication between components in the context of Policy Based Management, while ENTHRONE also designed special horizontal protocols like EQoS suite which is presented in Section 6. We should note that EQoS protocol is designed to hide heterogeneous IMS middleware implementation. It provides real-time communication, session identification, and state management functionality. Improved features of security, fault-tolerance, and message compression for bandwidth optimisation can be integrated easily.

Note that EQoS protocols related modules are belonging to the application layer, so that in principle, if desired, they can run as upper layer upon a standard DOT technology.

4. Multimedia content management

4.1. MPEG-21 metadata

MPEG-21 specifies an abstract framework for a multimedia delivery chain, and as such provides mechanisms and tools to declare, use, modify, metadata, and content. The fundamental unit of transaction in MPEG-21 is the Digital Item (DI) [13] that is a structured object with a standard representation.

A DI contains multimedia content (resources in MPEG terminology), associated metadata, and structures representing relation between resources and metadata. REL (Right Expression Language) [14] in MPEG-21 is used to describe the permitted right associated with the DI. The DIA metadata is used to facilitate the adaptation of the DI.

4.2. Metadata management at the IMS dispatcher

In order to provide users the transparent access to multimedia content with quality of service assurances, the IMS dispatcher will use descriptive metadata describing the multimedia content itself and contextual metadata, providing descriptions concerning the content format and possible operations to be performed for the necessary adaptation based on the usage rights, the network, the terminal, the usage environment characteristics, etc. While the content-descriptive metadata is fully addressed by MPEG-7and TV-Anytime [15] among other standards, contextual metadata is now being addressed by MPEG-21. In particular part 7 of the MPEG-21 standard – Digital Item Adaptation

(DIA) [16] – specifies a set of tools clustered into eight major categories that enable to adapt content in order to provide to the user, the best possible experience under the available conditions. This includes respecting user preferences, terminal capabilities, network characteristics, conditions, natural environment characteristics, etc.

4.3. Metadata integration by the IMS dispatcher

We can synthesize the integrated management of MPEG-21 metadata used in the our system with highlighting two distinctions: on one hand, from the dynamic perspective, metadata that are dynamically conveyed along with the digital content stream and metadata that are statically stored in local file systems or database management systems, and on the other hand, from the contextual perspective, metadata that are content-dependent (or independent) or context-dependent. Fig. 2 summarizes metadata framework being used in our architecture.

Content-dependant metadata such as Dublin Core metadata elements [17], MPEG-7 descriptors, TV-AnyTime metadata and MPEG-21 DIA and IPMP (Intellectual Property Management Protocol) information can be conveyed along with the digital content in the bit stream as shown in Fig. 2.

Other content-independent metadata can also be added for enabling the processing of DIA and IPMP information; they include MPEG-21 DIA and IPMP tool lists and containers that are the necessary and specific programs for the adaptation and the intellectual property management and protection of the DIs. Conjointly, additional context-dependent metadata are collected through the participating network: they include network QoS metadata, terminal metadata, usage, and environment metadata in MPEG-21 format. They are interpreted by the IMS dispatcher and sent to the DIA engine of the IMS for the adaptation processing. Context-dependent metadata comes from the network and the terminal device. The DI manager is a module of IMS middleware and it includes all the operations related to the processing of MPEG-21 information.

The format of metadata is XML encoded with BiM (XML Binary Mode). The bit stream is divided up into several sections of two types: the first type of sections of the bit stream is dedicated to the processing including the specific programs to use for adaptation or IPMP, and the second type of sections of the bit stream is dedicated to the content (audio, video content, and content-dependent metadata (among MPEG-7, TVA-AnyTime, Dublin Core elements, etc.). The IMS dispatcher receives the data and associated MPEG-21 metadata and sends Audio ES (Elementary Stream) and Video ES to the appropriate decoders. The message tool router manager of the IMS dispatcher sends the DI, the DIA metadata, the REL metadata, the IPMP metadata and the content-descriptive metadata to the appropriate engines and managers available in the ENTHRONE subsystems.

4.4. Metadata and QoS management at terminals

The end-user should be able to deal with media content in a transparent way in terms of different technologies (terminals, networks, content production mechanisms) and related options. The end-user Terminal should support the description of resources in terms of perceived quality and associated processing complexity. The TDM at the end-user terminal is designed to be able to:

- connect over a range of different access networks;
- describe terminal capabilities and access rights to the content distribution system and to the IMS middleware for QoS management in the framework of MPEG-21;
- issue requests for searching content available in the distribution system and issue requests for services;

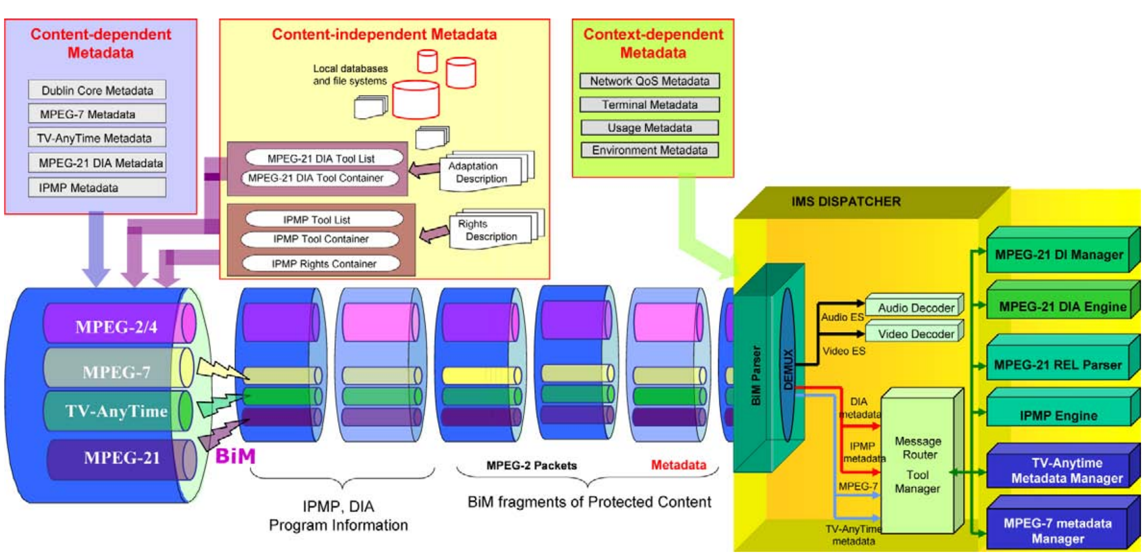

Fig. 2. MPEG-21 metadata managed by the IMS dispatcher.

- process descriptive and declarative information received with the content using MPEG-21 Digital Item Declaration and Digital Item Adaptation models. This information can be used to control the access to the data, to define the operations associated to the protected data, to define the adaptation operations specified by the IMS for QoS management;
- process the content according to the descriptive and declarative data provided by the Digital Item Adaptation model using the Digital Item Processing methods of the terminal taking into account the terminal capabilities and the achievable QoS provided by the IMS.

The TDM deals with efficient graceful degradation techniques supported by an MPEG-21 terminal middleware to satisfy decoding complexity and bandwidth requirements.

4.5. Content and metadata generation

Any basic activity related to content treatment (encoding, filtering, multiplexing, remultiplexing, content searching, etc.) is provided by its associated TVM (TV/Multimedia) Processor.

The TVM (TV/Multimedia) Processor is the main **ENTHRONE** content generation building block. All content processing and adaptation functionalities are handled by the TVM processors. The TVM takes as inputs, digital TV and/or Multimedia streams, then it adapts the content to the target format, according to the IMS requests. The generated stream format is determined by the end-to-end QoS mechanism decision as implemented in the IMS Dispatcher. To provide this functionality, the TVM processor is fully managed and controlled by IMS, via MPEG-21 management and control interfaces. Fig. 3 illustrates the interfaces of the TVM.

As can be seen, the TVM is located at the ingress of the distribution network (e.g., content servers). The TVMs are responsible for creating the processed content stream according to IMS configuration, and stream it over the network towards the receiving terminal.

Digital Rights Management (DRM) modules interact with the TVM to allow encryption of the outgoing stream. Some of the encryption modules are implemented as an integral part of the TVM, and some are performed externally.

Another task of TVM is to generate the content-descriptive metadata, in conformance with the MPEG-21 framework. The TVMs will generate or extract content-descriptive metadata in XML format (MPEG-7, TV-Anytime or MPEG-21 compliant), according to the processed content.

5. End-to-end QoS provisioning over Heterogeneous Networks

The end-to-end QoS provisioning issues in the project are tackled at many levels; at the terminal level, the network level and the content level. Fig. 4 shows the overall networking configuration considered in the end-to-end QoS chain.

The proposed solution for multi-domain QoS offering over heterogeneous networks include a set functionalities among which are: (1) QoS-based service management for service provisioning and offering, (2) QoS-based resource management (resource allocation techniques and network configuration functions that allows the establishment of the end-to-end QoS-enabled path, and (3) QoS-based resource and service monitoring that monitors any degradation in the selected path and offered services at customer or provider levels. It is viewed that the above functional-

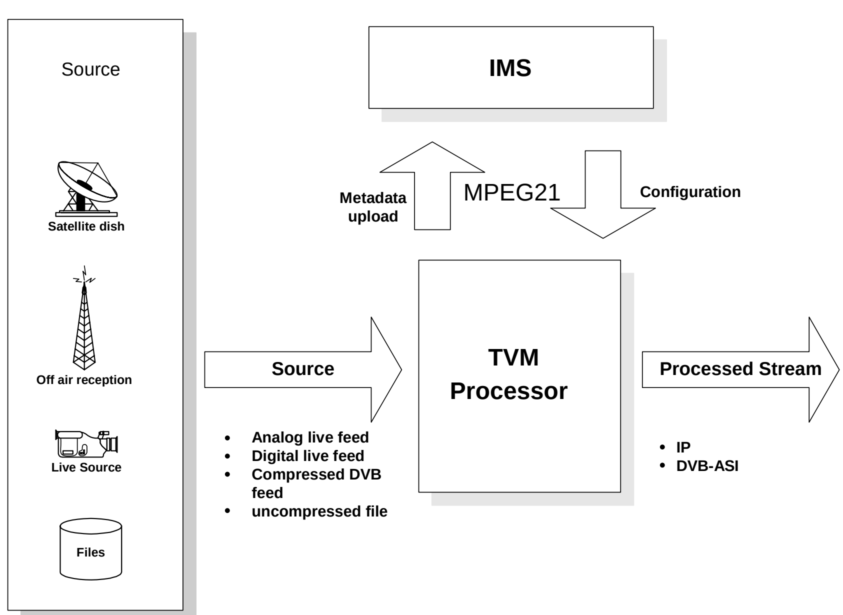

Fig. 3. ENTHRONE TVM processor.

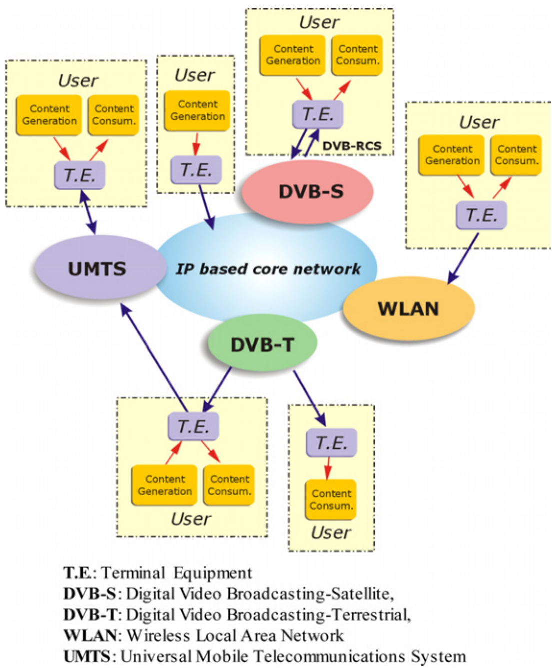

Fig. 4. Overall heterogeneous network configuration.

ities are realized through a suite of end-to-end signalling protocols, namely the **EQoS** (End-to-End QoS Signalling Protocol) [18]. These protocols provide a unified service/management plane governed by the IMS, whilst allowing network technology specifics within the domain scope.

In order to construct end-to-end QoS classes across multiple domains, it is envisaged that each domain is engineered to support a number of local QoS classes. That is, within a domain, all networking elements and links, and border routers/gateways are engineered and provisioned for offering these local QoS classes. These local QoS classes are mapped to globally well-known QoS classes, each for supporting one/more specific applications. The mapping and binding of these domain specific QoS classes are performed in order to offer end-to-end QoS classes across multiple domains. The characteristic of each well-known service is relying on common application QoS requirements and will be based on a common set of characteristics of transport mechanisms. A number of different classes for such set of characteristics would then exist, each with differing overall quality.

End-users use the same kind of applications in quite similar business contexts. Similar QoS requirements are likely to be requested by customers from their respective service providers. Providers are likely to define and deploy similar QoS classes locally, each of them being particularly designed to support applications of the same kind and with similar QoS performance constraints. These QoS classes can be extended end-to-end with mapping/binding of the similar local QoS classes achieved across the domains. Each local QoS Class is extended end-to-end along the QoS class characteristics but to meet end-to-end requirement of some targeted applications. The related end-to-end performance values should be bounded in order not to exceed the target values suitable for the targeted application services. This means that the scope of the end-to-end QoS class may only be broadened within the application's end-to-end performance boundaries and constraints.

5.1. QoS level peering arrangements

There must be service level interactions between SP and NPs and between NPs for constructing QoS-based services across multiple domains. A number of models is envisaged for QoS-based peering between service/network providers for providing services across multiple domains. These models are strongly influenced by experience in the telecommunications industry for the provision of international telephony and other services for which network interconnection is a requirement, both in commercial and regulatory terms.

The type of inter-domain peering impacts the service negotiation procedures, the required signalling protocols, the QoS class mapping/binding and path selection [19]. Fig. 5 shows two peering models:

- The **hub model** where the SP, as a distinct entity from NP, is the central point that negotiates and establishes provider-level SLSs with all individual NPs to be involved in the service chain.
- The **cascaded/bilateral model** where a SP and subsequent NP/s only negotiates provider-level SLSs with its immediate neighbouring provider(s) to construct an end-to-end QoS service. That is, only neighbouring domains establish provider-level SLSs between themselves.

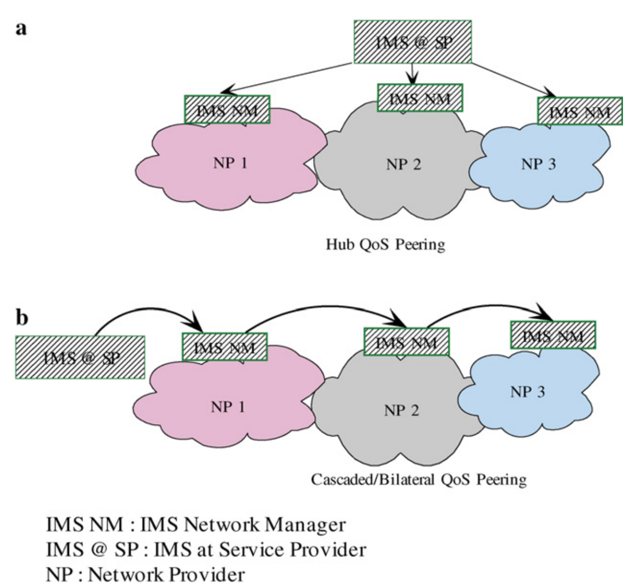

Fig. 5. QoS peering models. (a) Hub QoS peering. (b) Cascaded/Bilateral QoS peering.

- ENTHRONE adopted the cascaded model, which is more scalable, [19] since (1): the SP does not need to have/maintain inter-domain topology knowledge and (2) the total number of signalling messages between NPs can be decreased (in comparison to the Hub model) because an NP can aggregate in a single pSLS several requests that follows the same route towards the same destination. in hub model, the order of magnitude of the number of pSLSs needed for an NP to reach every other domain is $O(N^2)$ per QoS class, where N is the number of domains [19]. In the cascade model each NP has at most N-1 pSLS contracts to establish, with its neighbours in order to reach to N-1 destination domains. Therefore, we have $O(N)$ complexity for pSLS set-up per QoS class.

It should be noted that each NP can have its own network management system managing its domain in terms of resource provisioning which is de-coupled from the IMS-NM but should have the means to communicate with the IMS.

6. Overall functional architecture

We adopted a four-plane based architecture as described below for provisioning end-to-end QoS over multiple networking domains/autonomous systems.

Fig. 6 depicts the functional entities involved in the end-to-end QoS provisioning over heterogeneous networks. In particular it describes the deployed IMS middleware at NP.

6.1. Physical infrastructure

The Physical Plane includes the actual networking elements (e.g., terminals, servers, routers, switches, QoS capable multiplexers/encapsulators, etc.) and inter-domain links that form the networking infrastructure. These elements are grouped in domains, administered by a separate NP. Each autonomous system encompasses a specific networking technology (IP, DVB, etc.).

6.2. Data and control planes

Data-plane implements the binding functions from well-known QoS classes to local QoS classes and vise versa at the ingress and egress points of a domain. Several components in this plane are responsible for packet classification, marking, policing and traffic shaping according to the agreed SLSs. In IP networks, traffic prioritisation is achieved using appropriate DiffServ Code Point (DSCP) and queuing and scheduling mechanisms are realized using different PHBs with the appropriate configuration. In DVB networks, other technologies to support QoS control such as bandwidth management is employed.

A one-to-one logical abstraction for all the networking elements of the physical plane using the corresponding PEPs (Policy Enforcement Points) is used. Appropriate protocols (such as: SNMP, COPS, or proprietary ones) are used to control these elements.

The PEPs can be **ePEPs** (Edge PEPs) when they control edge devices or **cPEPs** (Core PEPs) when they refer to core network devices. The ePEPs and cPEPs have different

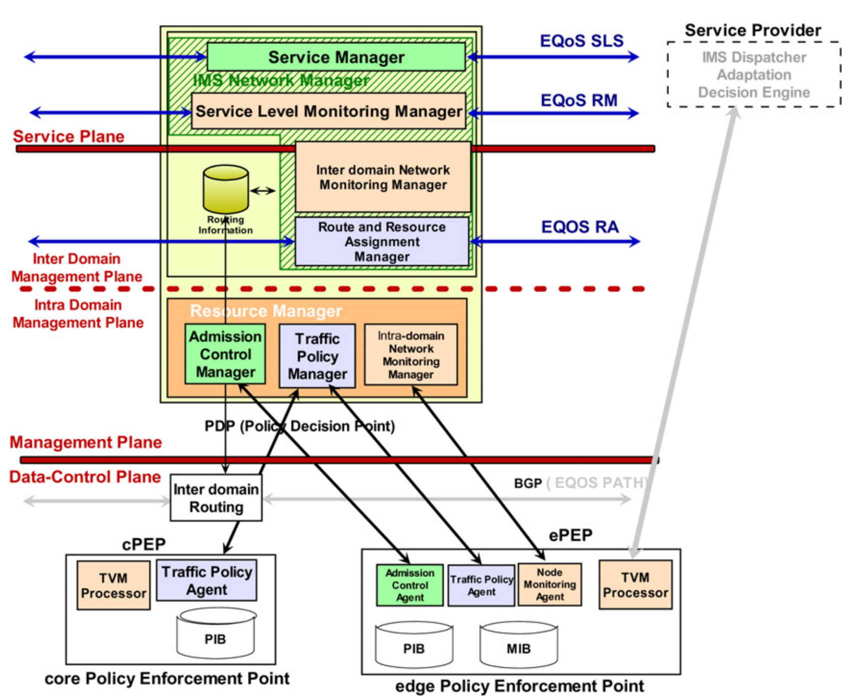

Fig. 6. High level functional architecture.

functionality. ePEPs include appropriate agents for call admission (i.e., Admission Control Agent), for traffic control (policing and shaping) according the Diffserv architecture (i.e., Traffic Policy Agent) and for QoS monitoring (i.e., Monitoring Agent). cPEPs contain only Traffic Policy Agents.

In IP-based networks, the classical BGP as the control plane protocol is essential for inter-domain routing. This may need to be extended for QoS-based route discovery and assignment.

6.3. Management Plane

The Management Plane is divided into two sub-planes: the intra-domain management sub-plane and the inter-domain management sub-plane. The first one is responsible for the management coordination of all the PEPs that control a single domain. The second one is responsible for the communication of the management functionality among the heterogeneous autonomous domains. The management plane is implemented through PDPs (Policy Decision Points) and there is a one-to-one relationship among PDPs and autonomous systems. A PDP is separated into two administrative modules, namely: (1) a Resource Manager (RM) for intra-domain provioning and configurations and (2) a network manager (NM) for handling inter-domain QoS provioning and configurations. It should be noted that the network configurations through PDP and PEP is considered when the distributed IMS is responsible for the network provisioning and allocation at both intra- and inter-domain levels.

The IMS-NM provides a homogeneous end-to-end view of the service for multiple domain heterogeneous networking. In general, the interface between the NM and RM defines the exchanged functionality between the IMS and domain-specific technology. This open interface is currently being standardized under the MPEG-21 Working Group. Note that an autonomous system that participates in the ENTHRONE environment may implement its own resource management functionalities. The IMS NM has to communicate directly with this module through the appropriate interface. The RM manages the PEPs of its own domain. It may contain functional element such Admission Control Manager, Traffic Policy Manager, and Intra-domain Network Monitoring Manager (as shown in Fig. 6. The Admission Control Manager sets up and manages the admission decisions of a particular traffic stream. The Traffic Policy Manager provides an automatic means for network administration and configuration. Intra-domain Network Monitoring Manager is responsible for composing monitoring information and making it available to the other providers. All these modules correspond to the respective agents of the corresponding PEPs.

The communication between RM and ePEP/cPEPs can be performed through a set of signalling protocols based on SNMP, COPS, or CLI.

6.4. Service Plane

In general, the Service Plane is responsible for the establishment of the SLAs/SLSs among the providers and/or customers that participate in the ENTHRONE infrastructure. In addition, it performs service planning and provisioning, QoS capability discovery, etc. at network provider level. The appropriate EQoS signalling protocol (**EQoS SLS**) is used for SLA/SLS negotiation.

6.5. EQoS-protocol suite

We make a distinction between service subscription through a SLA/SLS negotiation protocol and service invocation through a QoS-signalling or reservation protocol. Specifically, we view that an SLA negotiation protocol is used at the time of subscription for agreeing on QoS-based SLAs, whereas, QoS-signalling or reservation protocol is used for committing the resources by the provider when the user call for the service. The distinction between service subscription and invocation is required for checking conformance of user service requests against agreed profiles, which is essential in SLA-based service provisioning.

We developed a protocol for SLA/SLS negotiation (EQoS SLA/SLS), which is part of the EQoS Protocol suite [18]. A service can be established end-to-end whenever the resources are provisioned and allocated accordingly. The resource allocation allows a particular stream or streams to consume particular network resources based on the contracted SLS. The EQoS protocol suite includes a resource allocation protocol (EQoS RA).

QoS-based service and resource monitoring is essential to detect any service degradation and violations of the SLS agreements. A remedial function can be triggered to minimize the effect of service degradation. Such a function could be resource path renegotiations, path switching, traffic adaptation, etc. When traffic flows in a network, its actual resource use is monitored through the distributed monitoring system entities and this monitoring may be shared with the IMS entity at the SP. The EQoS protocol suite includes a resource monitoring signalling protocol (EQoS RM) to exchange monitoring information between domains. The whole monitoring architecture has been deeply investigated in this project. Implementation's results can be found in [20].

6.5.1. QoS-based signalling protocol for SLS negotiation and subscription

The customer formulates its request for service subscription and communicates with SP through SP-FE (Front-end). EQoS-SLA is used for this type of negotiation. The provider level agreement (pSLS) is set-up between SP and an NP or between two NPs as a commitment on resource usage for user traffic. This allows NPs to dimension and provision their network appropriately and to allocate their resources according to the requests and future demands.

Resource provisioning and allocation can happen either off-line during the network provisioning cycles or may be adjusted dynamically on demand. Reservation requests should not occur frequently as they may disrupt the services currently being provided. The developed provider level service negotiation protocol can provide simple request/response mechanism with a set of messages generically presented here as follows:

EQoS SLS request. This message is the customer/provider request for a service based on an SLS (cSLS, pSLS) from a provider.

EQoS SLS response. This message indicates whether the requested SLS can be fulfilled or not. This may trigger further negotiation. EQoS SLS Request and Response happens in horizontal direction.

Fig. 7 shows a simplified sequence of pSLS agreements between providers and cSLA agreement between customer and SP.

The developed signalling/negotiation protocol is more complex, as it covers both SLA and SLS negotiation allowing not only request/response dialogues as described above, but also handling more complex negotiation procedures.

With respect to negotiation features offered, the EQoS-SLA/SLS-NP is comparable to other negotiation oriented protocols like SrNP, [21], or COPS-SLS [22] but it is more adapted to the IMS system context and needs.

The EQoS-SLA/SLS-NP is a client-server half-duplex negotiation protocol between two negotiation entities. It functionality includes: setup a new, modify, or delete an SLA/SLS contract. The basic two-steps negotiation session is of type "proposal followed by yes/no_answer", started with a client proposal and then, server accepting or rejecting it. More complex several steps negotiation session can start with an initial proposal from the client, revision returned by the server, another proposal from the client and so on, up to the termination of the negotiation. Several variants (alternatives) of negotiated objects values and selective acceptance or rejection of different alternatives are possible. The conclusion of contract can be optionally delayed (negotiation session, started, interrupted, and resumed later). This feature could be useful for requests arriving sequentially to the server when the server is not able to serve the request immediately. More details are presented in [12].

6.5.2. QoS-based resource allocation protocol

A pSLS chain is established during the pSLS subscription for specific end-to-end QoS class. This chain can be from the point where the contents are located to a destination domain where an appropriate number of customers are located. The scope can be from the source domain (e.g., a router interface) to the ingress/egress point of the destination domain. Once, the pSLS chain is established and the resources are promised by the appropriate domains and during the pSLS invocation phase, the EQoS-RA protocol is run across domains in cascaded fashion in order to commit for allocation of resources. The EQoS-RA is a **data-path decoupled**, **out-of band** signalling protocol, running between IMS NM locat-

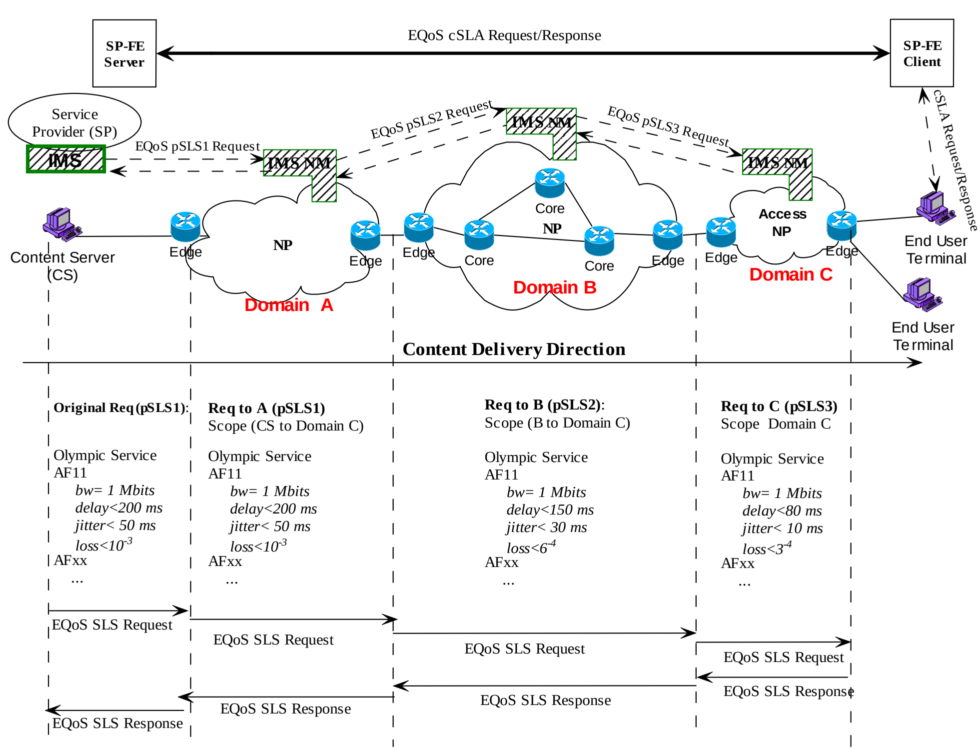

Fig. 7. SLS signalling protocol procedures.

ed at SP and NPs, i.e., it is not like RSVP protocol. The inter-domain resource allocation is accomplished at the pSLS aggregated level. This approach is important because it assures the scalability of the solution, by avoiding per-call signalling at inter-domain level. Some generic messages of this protocol can be:

EQoS RA reservation request. This message is from upstream domain to downstream domain for requesting allocation of the promised resources, (or a percentage of them).

EQoS RA response. This message is a response of the request message to indicate success/failure of the allocation. This message is sent from downstream to upstream domain and finally to SP.

The actual allocation implies vertical signalling actions between the domain Resource Manager (RM) to networking devices if necessary in order to install the appropriate configuration in Data Plane DiffServ entities (policers, meters, buffer management, and schedulers) in IP environment and multiplexers in DVB environment. Furthermore, delivering QoS guarantees requires controlling the amount of traffic entering the network. Therefore, appropriate information regarding the SLSs is given to be used by the domains for admission control and traffic conditioning purposes.

When the pSLSs are realized and their resources are allocated, then the SP can accept cSLA from customers, for individual subscription to some selected services (from a list offered by SP) with a given level of QoS. This means that SP deals with any cSLA requests and based on its prior knowledge of pSLSs installed it can accept or reject the requests.

At individual service invocation time the SP establishes a cSLS with the first NP situated at the entrance of the pSLS pipe (traffic trunk), in order for allowing the individual customer traffic to use the resources. The resources are already in place in the form of traffic trunk.

No signalling for this service invocation is needed between the NPs but only between CC and SP, CP and SP, and CP and CM, i.e., its content manger.

Note that each IMS distributed entity may aggregate several allocation requests in a single message. This can dramatically reduces the amount of requests between two domain peers and provides more scalability.

6.5.3. QoS-based resource monitoring protocol

Monitoring information can be used for network provisioning, dynamic resource allocation, route management, and in-service performance verification for value-added IP services. Here, a Resource Monitor always runs in each domain for performance and throughput measurements. The user traffic can be adapted during periods of congestion. If a domain fails to meet its contractual obligation, the IMS instructs the source to take remedial action or may apply a local adaptation. All these signalling and monitoring information are carried by a set of messages described in EQoS RM signalling protocol, which is currently being implemented. However, this inter-domain level monitoring is subject to establishing appropriate agreements between providers that allow a provider to retrieve monitoring information form other provider's respective domain/s.

7. Conclusion

In this paper, we have described a distributed management approach for providing end-to-end quality of service over heterogeneous networks. This system covers an entire audio–visual service distribution chain, including content generation and protection, distribution across QoS-enabled heterogeneous networks and delivery of content at user terminals. The whole content delivery chain is considered, including content, network, and service providers. We focused our attention on service provisioning, multimedia content management, and QoS provisioning over heterogeneous networks. We develop a customer/end-user service framework, defining and offering services through Service Providers. Service provisioning and service offering are achieved in a scalable manner. Both IP and non-IP networks are considered for resource management and configurations at the user terminal are realised using MPEG-21 features. In providing a QoS-enabled inter-domain networking infrastructure, a QoS-based signalling protocol suite is proposed for service subscription, resource allocation and monitoring purposes. The dialog between different stakeholders in the content delivery chain is considered as an important factor and a protocol is developed for these negotiations. Service and resource monitoring is essential for service providers. Means are developed to exchange measurement information and monitor the services, networks, and resources. In summary, we believe that the presented approach is scalable and can support the QoS-based service offering to a large number of customers across heterogeneous environments.

Finally, we have completed the design stage in which the proposed functional blocks of the architecture have been specified in terms of interfaces to other blocks, behavior and algorithms. We are currently at the stage of implementation and validation. We will experiment the system through both testbed environments using Linux-based routers and simulators in order to be able to deal with large-scale networks, stress conditions, etc.

Acknowledgements

This work was undertaken in the Information Society Technologies (IST) FP6 ENTHRONE project, which is partially funded by the Commission of the European Union. The authors would like to thank their colleagues who participated in the ENTHRONE project for the fruitful discussions, which have also contributed to the ideas presented in this article.

References

[1] D. Goderis, et al., Service level specifications semantics, parameters, and negotiation requirements, IETF Internet-Draft, draft-tequila-sls-02.txt, February 2002.

[2] B. Quoitin, C. Pelsser, L. Swinnen, O. Bonaventure, S. Uhlig, Interdomain traffic engineering with BGP, IEEE Communications Magazine 41 (2003) 122–128.

[3] P. Pongpaibool, H.S. Kim, Providing end-to-end service level agreements across multiple ISP networks, Computer Networks 46 (2004) 3–18.

[4] European IST (Information Society Technologies) research projects, for more information visit: <www.cordis.lu/ist/>. Specifically for IST-AQUILA visit: <http://www-st.inf.tu-dresden.de/aquila/>, for IST-CADENUS visit: <http://wwwcadenus.fokus.fraunhofer.de/old_entry.html>, and for IST-TEQUILA visit: <http://www.ist-tequila.org/> and for IST-MESCAL visit: <http://www.mescal.org/>.

[5] S.P. Romano (Ed.), Resource Management in SLA Networks, D2.3 CADENUS Deliverable, May 2003.

[6] G. Cortese, P. Cremonese, S.D'. Antonio, M. Esposito, S.P. Romano, A. Diaconescu, CADENUS: creation and deployment of End-User Services in Premium IP Networks, IEEE Communications Magazine (2003) 54–60.

[7] P. Trimintzios, G. Pavlou, P. Flegkas, P. Georgatsos, A. Asgari, E. Mykoniati, Service-driven traffic engineering for intradomain quality of service management, IEEE Network Magazine 17 (3) (2003).

[8] E. Mykoniati, C. Charalampous, P. Georgatsos, T. Damilatis, D. Goderis, P. Trimintzios, G. Pavlou, D. Griffin, Admision Control for Providing QoS in IP DiffServ Network: the TEQUILA Approch, IEEE Communications, special issue on QoS Advances: the European IST Premium IP Projects, IEEE 41 (1) (2003) 38–44.

[9] A. Asgari, R. Egan, P. Trimintzios, G. Pavlou, Scalable monitoring support for resource management and service assurance, IEEE Network Magazine (2004).

[10] P. Flegkas, et al., Mescal Deliverable D1.1, Specification of Business Models and a Functional Architecture for Inter-domain QoS Delivery, <http://www.mescal.org/>, June 2003.

[11] EUQOS IST project web site: <http://www.euqos.org>.

[12] Hamid Asgari (Ed.), ENTHRONE Deliverable D24i, Specification of protocols, algorithm, and components, the architecture, and design of SLS Management, October 2004, <http://www.enthrone.org>.

[13] ISO/IEC DTR 21000-1 MPEG-21 Vision, Technologies and Strategy, version 2, April 2004.

[14] I SO/IEC FCD 21000-5 MPEG-21 Rights Expression Language, October 2002.

[15] ISO/IEC 15938-1 MPEG-7 System – March 2001, ISO/IEC 15938-5 MPEG-7 Multimedia Description Schemes – March 2001, ISO/IEC 15938-5 MPEG-7 Multimedia Description Schemes Extensions – October 2002, ISO/IEC 15938-1:2001: Multimedia Content Description Interface (MPEG-7) – Part 1: Systems.

[16] ISO/IEC FCD 21000-6 MPEG-21 Rights Expression Language, October 2002.ISO/IEC FCD 21000-7 MPEG-21 Digital Item Adaptation, July 2003.

[17] Dublin Core Metadata for Resource Discovery. Internet RFC 2413. <http://www.ietf.org/rfc/rfc2413.txt>.

[18] Toufik Ahmed (Ed.), ENTHRONE Deliverable D23i, End-to-end QoS Signalling and Policy-based Management Architectures, October 2004, <http://www.enthrone.org>.

[19] Hamid Asgari (Ed.), Mescal Deliverable D1.4, Issues in MESCAL Inter-Domain QoS Delivery: Technologies, Bi-directionality, Interoperability, and Financial Settlements, January 2004, <http://www.mescal.org>.

[20] A. Mehaoua, T. Ahmed, H. Asgari, M. Sidibe, A. Nafaa, G. Kormentzas, T. Kourtis, Service-driven inter-domain QoS Monitoring System for Large-scale IP and DVB Networks, to appear in Computer Communication Journal, SI on Monitoring and Measurements of IP Networks, 2005.

[21] T. Damilatis, et al., Final Architecture, Protocol and Algorithm Specification, TEQUILA IST Project, Public Deliverable, October 2002.

[22] Thi Mai Trang Nguyen, Nadia Boukhatem, Guy Pujolle, COPS-SLS usage for dynamic policy-based QoS Management over heterogeneous IP Networks, IEEE Network (2003) 44–50.

Toufik Ahmed received his Master's and Ph.D. degrees in computer science from the University of Versailles in 2000 and 2003, respectively. He was a visiting scientist at the School of Computer Science of the University of Waterloo in 2002 and research fellow at PRiSM laboratory of the University of Versailles in 2003. He is currently an associate professor at ENSEIRB school of engineers/University of Bordeaux I. His main research activities concern Quality of Service for multimedia wired and wireless networks, end-to-end signalling protocols, and policy-based networking.

Abolghasem (Hamid) Asgari (Hamid.Asgari@thalesgroup.com) received his B.Sc. from Dr. Beheshti University in Tehran, M.Sc. (Honors) from the University of Auckland, New Zealand both in Computer Science, and his Ph.D. in the design and analysis of ATM networks from University of Wales, Swansea in 1997. He was with Iran Telecom. Research Center (ITRC) from 1986 to 1990. He joined Thales (formerly Racal) Research and Technology (UK) in 1996 where he is an Assistant Chief Engineer/Project Manger specialising in the data communication systems/networks. He was involved in simulation, design and analysis of integrated IP services and network architectures. He has worked on a number of European Commission projects including IST TEQUILA, MESCAL and ENTHRONE projects. His main research interests are in network QoS management, QoS-aware monitoring, network and system level performance evaluation, traffic engineering, and service management.

Ahmed Mehaoua is an associate professor on Telecommunications in the Computer Science department at the University of Versailles, France. He is currently the leader of the Multimedia and Mobile Networking Group at the CNRS PRISM laboratory, a governmental computer science research center. Prior to join Univ. of Versailles, he was with University of Cambridge, UK, as an associate research fellow in the Center for Communications Systems Research, a join Cambridge Univ/Hitachi IP networks. His research has been supported by funds from government and corporate organizations (European Commission, French Ministry of Education and Research, CNRS, THALES, France Telecom, Ericsson).

Laure Berti-Equille is currently Associate Professor at the University of Rennes (France). Her research interests at IRISA lab (INRIA-Rennes, France) are multi-source data quality, quality-aware data integration, and multimedia metadata and multimedia mining. In the ENTHRONE project, she is responsible for metadata specifications and multimedia metadata authoring tools.

George Kormentzas is currently lecturer in the University of the Aegean, Department of Information and Communication Systems Engineering. He was born in Athens, Greece on 1973. He received the Diploma in Electrical and Computer Engineering and the Ph.D. in Computer Science both from the National Technical University of Athens (NTUA), Greece, in 1995 and 2000, respectively. From 2000 to 2002, he was a research associate with the Institute of Informatics and Telecommunications of the Greek National Center for Scientific Research "Demokritos". His research interests are in the fields of traffic analysis, network control, resource management and quality of service in broadband networks. He has published extensively in the fields above, in international scientific journals, edited books and conference proceedings. He is a member of pronounced professional societies, an active reviewer and guest editor for several journals and conferences and EU-evaluator for Marie Curie Actions. George Kormentzas has participated in a number of national and international research projects, serving in some instances as the project's technical representative for University of Aegean and/or as WP leader and/or as the project's Technical Manager.

Eugen Borcoci received his MS and Ph.D in Telecommunications from University "Politehnica" of Bucharest and MS in Mathematics from University of Bucharest, Romania. From 1995, he is professor in the Telecommunications Department of Electronics and Telecommunications and Information Technology Faculty of University "Politehnica" of Bucharest. His current teaching and research areas of interest are IP based architecture for integrated networks and services, communication protocols, QoS issues, multimedia. He participated to many national and international research projects related to the above areas. Mr E. Borcoci is member of IEEE Communication Society.