A QoS Negotiation and Adaptation Framework for Multimedia Services in NGN

L. Skorin-Kapov *, M. Matijasevic†
*Ericsson Nikola Tesla, R&D Center, Zagreb, Croatia
†University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia
lea.skorin-kapov@ericsson.com, maja.matijasevic@fer.hr

***Abstract*— A key challenge for the next generation network (NGN) is securing Quality of Service (QoS) support for personalized and advanced multimedia services. In this paper we present a comprehensive and generic SErvice-Level End-to-end QoS negoTiation and adaptation (SELEQT) framework. The framework is comprised of four conceptual models addressing and relating the following aspects: the actors involved in QoS negotiation (Business Model); specification of the parameters that impact the QoS negotiation process (Data Specification Model); the identification of dynamic adaptation-triggering events that lead to QoS renegotiation and adaptation (Adaptation Event Model); and the process of matching and coordinating QoS parameters specified and signaled by involved actors (Negotiation Model). Framework applicability is illustrated using an example multimedia service scenario.**

I. Introduction

In the context of all-IP multiservice next generation networks (NGN) [1], there is a perpetual challenge to review and match multimedia and networking characteristics of advanced multimedia services to dynamically manage quality of service (QoS).

With the onset of new business models on the telecom market, flexible and extensible QoS negotiation and adaptation (QNA) mechanisms, applicable for existing and emerging end-to-end (E2E) service scenarios and contexts, may be considered key enablers for providing customized multimedia services. Related work addressing QNA has focused on various aspects of the problem including QoS signaling issues [5] and the description of various service and transport configurations used for service negotiation [6][7], service adaptation platforms based on profile matching [8], and decision-making for the optimization and adaptation of service parameters [9][10][11]. However, providing a comprehensive approach to solving service-level QNA in a converged NGN architecture uniting both technical and business aspects is still considered an open issue [19].

In this paper, we propose a new SErvice-Level E2E QoS negoTiation and adaptation (SELEQT) framework. The framework is generic in the sense that it is independent of a particular network scenario, service scenario, and applied technology. The framework identifies and relates different aspects and viewpoints of QNA in four interrelated models: i) the actors involved in QoS negotiation (*Business Model*); ii) the specification of the parameters that impact the negotiation process (*Data Specification Model*); iii) the identification of dynamic adaptation-triggering events that lead to QoS renegotiation and adaptation (*Adaptation Event Model*); and, iv) the process of matching and coordinating QoS parameters specified and signaled by involved actors (*Negotiation Model*). With our focus being on service-level functionality, the underlying mechanisms and architecture that provide actual service delivery and network-level QoS support are presumed, but not addressed further in the scope of this paper.

A key contribution of the framework is the addition of a *QoS Matching and Optimization Function* (Q-MOF) in the service control layer of the NGN architecture to be included along the E2E signaling path. The Q-MOF is designed to support optimized service delivery and controlled service adaptation in light of changing resource availability. Furthermore, what is missing in most current approaches is the specification and signalling of an optimal degradation path for multimedia services composed of multiple flows. Our solution is for the Q-MOF to calculate a *media degradation path* (MDP) that will be signaled to communication end points and to the network layer to aid in network resource (re)allocation.

The paper is organized as follows. In Section II we introduce the fundamental concepts and terminology used in this work. In Section III we describe the SELEQT framework and its four conceptual models. Section IV describes a possible implementation of the proposed functionality within the NGN architecture. Section V illustrates framework applicability using an example multimedia scenario. In Section VI we discuss contributions with respect to related work and summarize the conclusions. Section VII presents issues for current and future work.

II. Fundamental concepts

A multimedia service is a combination or a set of combinations of two or more *media components* (e.g., audio flows, video flows, graphics data, etc.). We assume that a multimedia service exists in one or more different versions to meet the heterogeneous capabilities of end users and access networks. We specify *service versions* as differing in the included media components (e.g., Version 1 with streaming media and Version 2 without streaming media). Each media component may be configured by choosing from a set of offered alternative *operating parameters* (e.g., different codecs, frame rates, resolutions, etc.). For a particular service version, we refer to the overall *service configuration* as the set of chosen operating parameters for all included media components.

Service adaptation may be defined as the change of actual service version and/or service configuration. Further on, we consider service adaptation to be based on E2E QoS negotiation for a particular session, the purpose of which is that all involved actors reach an agreement regarding service configuration, resource allocation, and the dynamics of service adaptation. An example of service adaptation involves a user switching to a lower bit rate video codec due to a decrease in bandwidth availability.

We illustrate the steps involved in QoS QNA in Figure 1 in order to introduce the concepts that will further be elaborated in the SELEQT framework. Upon a service request, relevant input data specified by involved actors are collected, and a *matching process* is invoked to determine a set of feasible service versions and operating parameters based on evaluation of the following: (1) user terminal, access network, and core network capabilities; (2) user requirements (e.g., in terms of acceptable cost, media components and timing constraints), and (3) service requirements. This set of feasible service versions and operating parameters is the *Feasible Service Profile* (FSP).

The subsequent step is the *negotiation process* in which the FSP is offered to relevant entities in order to achieve an E2E agreement: an end user may accept, refuse, or modify the offered feasible parameters, and network entities may further authorize resources based on offered feasible parameters. Negotiated and authorized parameters are input to the *optimization process*, which calculates the optimal service configuration and respective resource allocation for all media flows. The optimal service configuration, and a number of alternative (suboptimal but feasible) service configurations are ordered by descending utility and signaled to involved entities in the form of an *Agreed Service Profile* (ASP). The ordered list of alternative configurations within the ASP is referred to as a *media degradation path* (MDP).

III. SELEQT FRAMEWORK

In order to provide a comprehensive view of the functionality and processes necessary to support the steps illustrated in Figure 1, a generic framework is proposed. The SELEQT framework (Figure 2) is composed of four conceptual models addressing and relating different aspects. The *Business Model* serves to identify the actors involved in E2E QNA, and model their roles and relationships. The *Data Specification Model* provides a description of the parameters (specified by relevant actors) that impact the QoS negotiation process. The *Adaptation Event Model* models the various dynamic scenarios that lead to QoS renegotiation and adaptation. The *Business*, *Data Specification*, and *Adaptation Event Models* define

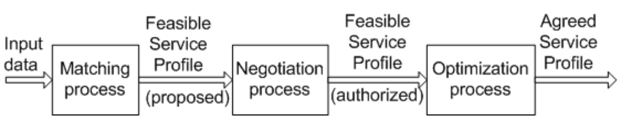

Figure 1. High-level view of steps in the QNA process

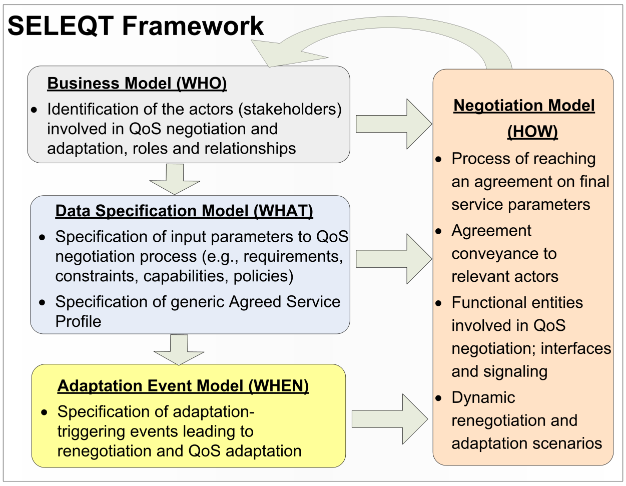

Figure 2. SELEQT framework

the environment that impacts the QoS *Negotiation Model*, which in turn defines the QoS negotiation process and identifies the functional entities that are involved in this process. For more details about SELEQT, the interested reader is referred to [12].

A. Business Model

The identification of actors involved in QoS negotiation is crucial in that actors dictate the parameters that need to be coordinated in order to deliver a customized E2E service. Two generic roles that may be defined are the *user* and *provider*, where a user is an entity that uses a service provided by a provider. The SELEQT framework specifies a generic Business Model as shown in Figure 3 (in the context of SELEQT, this refers to specification of actors, roles, and relationships and is a subset of what may generally be considered a business model in broader view). A concrete application of this model depends on the actors involved and the business and technical relationships established along the service delivery value chain. We assume the "one-stop responsibility" concept (adopted by the ITU-T in [13]) with the *primary service provider* being responsible for coordinating the QoS negotiation process, while further relying on the services of *sub-providers* in order to secure an E2E QoS.

An actor may take on a number of roles in a particular scenario, while a number of actors can play the same role. When observing the relationships between two actors, both business and technical aspects are considered. Two actors (e.g., user and service provider) may have an agreement specifying a business relationship (e.g., based on

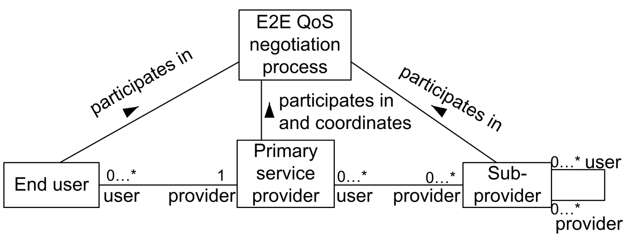

Figure 3. SELEQT framework: Business Model

payments), while technically actual traffic flows through a different actor's (e.g., network provider) domain.

B. Data Specification Model

A number of existing approaches in standards and literature deal with specifying various data sets that are relevant for negotiating QoS, such as: the specification of user capabilities (e.g., the Composite Capabilities/Preferences Profile (CC/PP), and in broader view the Mobile Web Initiative, MPEG-21 DIA [11]); service related data (IETF Session Description Protocol (SDP) [7] commonly used together with the Session Initiation Protocol (SIP) [5], End-to-End Negotiation Protocol (E2ENP) [6]); and operator policies [14]. What is missing is a high level data model to be used as a reference by multiple actors.

The goal of the Data Specification Model is to identify the parameters that serve as input to the negotiation process and the parameters that comprise the output. Key contributions of the Data Specification Model are given in our previous work [15], and are summarized in this paper for the sake of completeness. Figure 4 shows a generic model of the input data. Starting from a business model, the roles and relationships taken on by an actor will dictate the data that the actor will specify in the context of QoS negotiation. The input data consists of *associations*, which are defined to encompass logically related data. Typical associations include "User Data", "Service Provider Data", "Service/Application Data", and "Network Provider Data". Logically related data components within the association "User Data" include, for example, "user preferences", "access network capabilities", "terminal capabilities", and "subscription data".

The component in the lowest level contains one or more parameter sets, each composed of one or more parameter-value relations. For example, we can have an association "User data" which contains a component "user preferences", which in turn contains a parameter set "video preferences", which in turn contains the parameters "codec=MPEG", "frame rate=25", and "desired quality = HIGH". It is important to note that QoS parameters are specified at different abstraction levels (e.g., user perceived QoS, application level QoS, network QoS) and need to be mapped down to resource requirements (e.g., bandwidth determined based on codec output rate and frame rate).

Figure 5 shows an example of input data that needs to be coordinated during QoS negotiation, including a typical set of identified roles, corresponding associations, and different components. A more elaborated view of User Data and Service/Application Data modelled as a generic *User Profile* and a generic *Service Profile* respectively is given in [15].

In the scope of the generic *Service Profile* (specified by the actor that is providing the service to the end user in order to describe service characteristics), we base our method of specifying service requirements and adaptation capabilities for media flows on Utility Functions (UF). UFs may be adopted to specify the relationship between achievable quality and necessary resources [16].

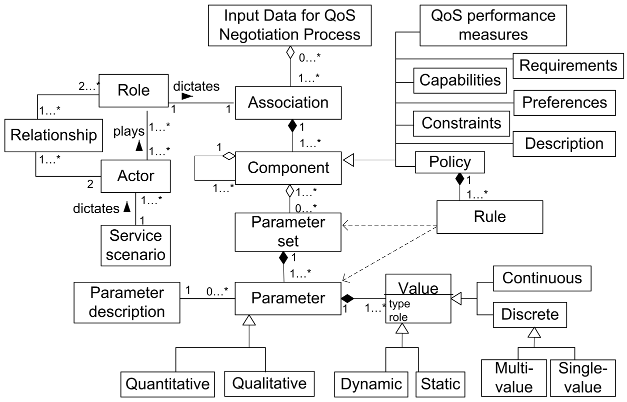

Figure 4. Model of input data specification for QoS negotiation

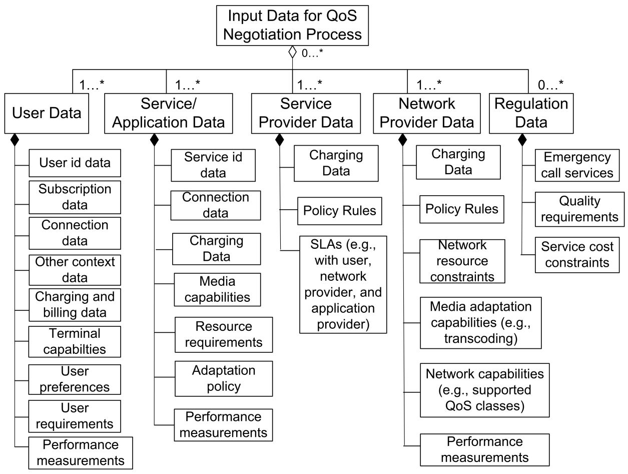

Figure 5. Example of input data (top three hierarchy levels portrayed)

The notion of a Utility-Based Adaptation Framework is discussed in [17] with the authors specifying a mapping between application adaptation space, resource constraints, and achieved utility. Mechanisms proposed in [17] and also [10] have been adopted in the MPEG-21 Digital Item Adaptation (DIA) standard [11] as part of the *AdaptationQoS* tool to be used for meta-data driven QoS adaptation. We adopt these mechanisms and apply them in the broader context of service-level (re)negotiation. For a given media component within the Service Profile, we define an Operating Space O whose dimensions constitute those service-level parameters that are being negotiated (e.g., codec, frame rate). We assume a mapping of a point in O to a Resource Space R indicating the (network and system) resources required to support that operating point. A point in O and R is further mapped to a Utility Space U, which may consist of multiple dimensions representing both qualitative and quantitative quality measures. In making a decision on how to optimally configure a given service, one operating point will be chosen for each media flow, and corresponding resources requirements will be used for resource allocation.

An Agreed Service Profile (ASP) represents the output of a successful QoS negotiation process as mentioned in Section II. It contains a set of (feasible) service versions ordered according to decreasing user perceived value,

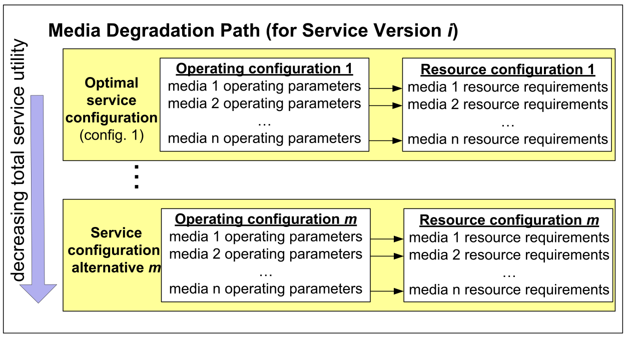

Figure 6. Logical view of a generic Media Degradation Path

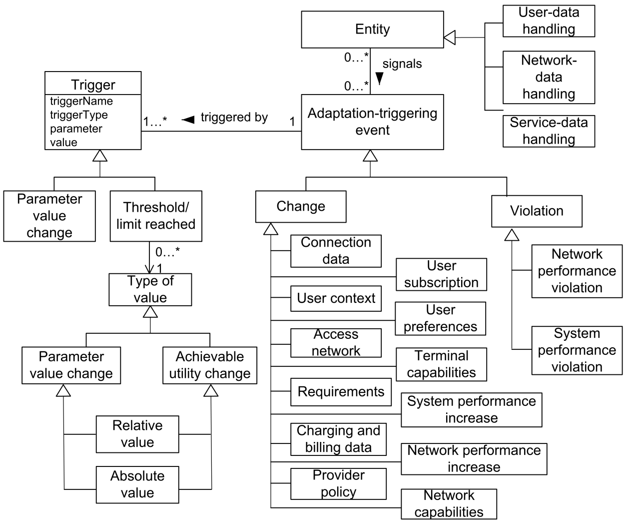

Figure 7. SELEQT framework: Adaptation Event Model

with the top ranked service version marked as *chosen service version*. For a chosen service version, an MDP is specified and included in the ASP. The MDP is signaled to the network level and interpreted as a list of possible alternative resource allocations in the case that the optimal resource allocation request cannot be achieved. The MDP may be used to steer service adaptation during the service lifetime based on changing resource availability. Such an approach would lead to quicker adaptation times, rather than having to go through the process of renegotiating session parameters with each change in resource availability. Figure 6 shows a logical view of an MDP with n different media components and m feasible service configurations for a particular service version i. It should be noted that by specifying the MDP we are considering the effect of degradation (or upgrading) simultaneously on all involved flows, rather than only for a single flow.

While the MDP specifies the logic of *how* a service should be adapted given a certain network and system resource allocation (by way of operating parameters), there remains the question of *when* it makes sense to adapt. What is missing in currently available standards (e.g., SIP/SDP) is how the involved entities (end systems, network nodes) will agree on the conditions (triggers) under which they will need to signal to each other that adaptation is required. We refer to such signals as adaptation-triggering events (ATE).

C. Adaptation Event Model

We specify the Adaptation Event Model shown in Figure 7 to model the various dynamic ATEs that may lead to QoS renegotiation and adaptation (referring to both service adaptation and modification of network/system resource allocation). The model portrays different entities signaling ATEs, including those responsible for handling user-data, network-data, or service-data. We define an ATE as a signal indicating that a violation has occurred with regards to agreed network/system performance (e.g., a decrease in the network bandwidth allocation past a given threshold), or that a change has occurred in the input parameters necessary for the QoS negotiation process (e.g., addition of new media stream). While certain service adaptation mechanisms (e.g., media scaling, frame dropping, etc.) may be invoked directly without the need to exchange signaling messages, we focus on those events that lead to signaling messages being sent between involved entities along the signaling path. Adaptation may be invoked by service logic or resource allocation mechanisms in response to a signaled ATE if corresponding service parameters have already been negotiated. Otherwise, it may be necessary to renegotiate service parameters prior to adaptation. ATEs may be signaled in various ways, depending on the protocol that is being used. In the case of SIP-based signaling, events may be signaled using messages such as UPDATE, re-INVITE, and NOTIFY.

An ATE may be triggered by one or more different triggers. Two different types of triggers have been identified: (1) *threshold/limit reached*, meaning a value has changed past a certain threshold/limit and requires some sort of action; and (2) *parameter value change*, meaning that the change in parameter value requires some sort of action.

The value of a limit/threshold trigger may be expressed in absolute or in relative terms with respect to either the parameter value, or, the achieved utility. Using as a basis the Adaptation Event Model, we introduce the notion of ATE filters to be included in the previously introduced ASP to specify the adaptation triggers for a particular service, hence limiting events signaled during the course of a session to those of relevance to the service (e.g., derived based on service adaptation capabilities specified in the Service Profile) and in line with provider policy (e.g., the provider wishes to reduce signaling traffic and therefore imposes restrictions on the sending of updates; the provider does not offer support for certain updates; a user's subscription does not support adaptation, etc.). It should be noted that, when signaling an ATE, there is no need to re-send all previously signaled profile parameters.

D. Negotiation Model

Within the context of the SELEQT framework, the Negotiation Model uses as a basis the specified Business, Data Specification, and Adaptation Models to specify the process of matching and coordinating relevant input parameters signaled by actors involved in the E2E QNA. Two key processes included in the QoS negotiation are the Matching Process and Optimization Process, shown

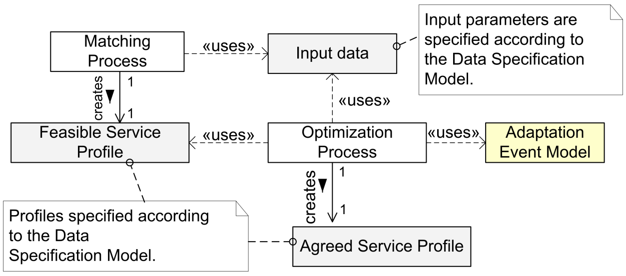

Figure 8. SELEQT framework: Negotiation Model

in Figure 8 in relationship to other related elements of the other three conceptual models.

The implementation of the framework in the context of an NGN architecture, with emphasis on these two processes, is described next.

IV. Implementation of Seleqt within NGN architecture

For the purpose of coordinating QoS negotiation, we introduce a novel network-based QoS Matching and Optimization Function (Q-MOF) located in an NGN Service Provider (SP) domain, responsible for collecting input data, conducting matching and optimization of parameters, and determining a final agreement among involved entities in the form of an ASP.

During the process of QoS (re)negotiation, signaling flows typically traverse a number of functional network entities along the E2E path. A QoS Negotiation Functional Architecture (QNFA) is proposed in Figure 9 which illustrates key generic functional entities involved in the negotiation process. The QNFA adheres to the NGN functional architecture as specified by the ITU-T [1] separating NGN functions into two layers: a service stratum and a transport stratum, with communication QoS parameters being negotiated at the call/session layer. We portray communication end points as either end users or application servers offering applications and services. We assume an end user as having a primary SP that performs service control functions and offers access to 3rd party applications and services. The SP further interacts with an underlying Network Provider (NP) domain (access network) that provides media connectivity functions (a single operator may take on multiple roles, including both a SP and NP). The SP domain may be accessed via multiple access networks. As shown in the figure, we propose for the Q-MOF to be included in the SP domain as a generic functionality, allowing for it to be reused by services requiring advanced QoS support. In addition to providing a better service to users, introducing enhanced QoS support in the network as a reusable service capability would benefit both the SP and third-party service/application providers. The SP would have additional means to control, differentiate, and appropriately charge the QoS a particular user receives for a given multimedia service. Third-party service/application providers would have to specify a service profile stating service requirements and options and would further be relieved from implementing complex QoS decision making functionality for each new service, hence leading to simplified provisioning and possibly quicker time-to-market for new services requiring such mechanisms.

A. *QoS Matching and Optimization Processes*

The aim of the Matching Process conducted by the Q-MOF is to parse collected input data (example input data shown in Figure 5) and match service requirements against the capabilities/requirements of service users, and any additional imposed constraints (e.g., imposed by network policy) in order to determine feasible service parameters. As output, the Feasible Service Profile (FSP) specifies all feasible service versions and operating parameters. After entities have agreed to the FSP, the FSP is passed on to the Q-MOF Optimization Process. If matching results indicate that requirements cannot be met, the Q-MOF will signal that no FSP can be found.

The Optimization Process calculates the optimal service configuration (operating point) and the corresponding network and system resource allocation, and as output produces the ASP. The optimization problem will be formulated using the following:

- the FSP specifying the Operating-Resource-Utility (O-R-U) mapping for active media flows;
- user preferences used to assign relative weight factors to resource allocations for different media flows;
- service costs (specified depending on the charging models being applied. Examples include: cost per unit time, cost per byte, flat rate, etc.).
- network/system resource constraints and bounds on values (e.g., bandwidth, user budget).

In searching for a solution, we assume a finite number of feasible choices or operating points per media flow. The goal is to choose exactly one operating point per media flow so as to maximize utility. For simplification purposes, we assume a one dimensional utility space corresponding to user perceived value. The general problem may be formulated as follows. We assume n different media flows, of which flows $1, \ldots, h$ are in the downlink direction and flows $h+1, \ldots, n$ are in the uplink direction. The ith flow has p_i operating points. Required resources $\mathbf{r}$ for operating point j and media flow i are denoted as $\mathbf{r}_{ij} = (r_{ij1}, \ldots, r_{ijq}, \ldots, r_{ijk})$.

Assuming q different QoS classes, we specify $r_{ij1}, \ldots, r_{ijq}$ as corresponding to bandwidth assigned within different QoS classes (differing in delay, loss, jitter, and bandwidth guarantees). For a single operating point, we assume only one of the values $r_{ij1}, \ldots, r_{ijq}$ to be greater than zero, while all others are equal to zero. This is because only one QoS class is chosen per operating point for a media flow. We define variables $B_{downlink}$ and B_{uplink} to denote maximum available downlink and uplink bandwidth respectively. They represent the resource constraints for $r_{ij1}, \ldots, r_{ijq}$. Additional resource constraints (other than bandwidth) related to resource consumption across all media flows are expressed as $\mathbf{R} = (R_{q+1}, \ldots, R_k)$ (e.g., cost for all flows must be less than a specified amount). Resource constraints may also be added regarding resource consumption per flow

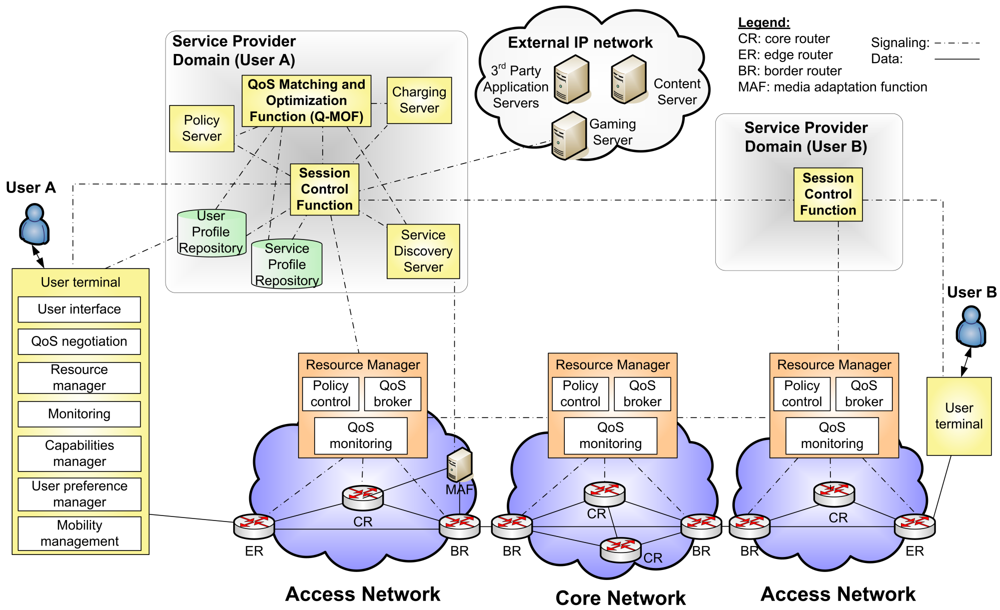

Figure 9. QoS Negotiation Functional Architecture

and expressed as $\mathbf{R}_i = (R_{i1}, \ldots, R_{ik})$ (e.g., cost for a media flow must be less then a specified amount).

The utility value for operating point j and media flow i is denoted as $u_i(\mathbf{r}_{ij})$. Weight factors w_i are assigned to utility values to indicate the relative importance of media flows (we note that the process of assigning weight factors to media flows is out of scope for this paper). We include 0-1 variables x_{ij} to make sure that exactly one operating point is chosen per media flow.

Below we show formulation for the case of multi-flow resource constrained utility maximization, where in calculating optimal resource distribution among flows, we consider total utility to be the weighted sum of individual media flow utilities. The problem is formulated as a multi-choice multi-dimension 0-1 knapsack problem (MMKP).

$$max \sum_{i=1}^{n} \sum_{j=1}^{p_i} w_i x_{ij} u_i(\mathbf{r}_{ij}) \tag{1}$$

such that

$$\sum_{i=1}^{k} \sum_{j=1}^{p_i} \sum_{y=1}^{q} x_{ij} r_{ijy} \leq B_{downlink} \tag{2}$$

$$\sum_{i=k+1}^{n} \sum_{j=1}^{p_i} \sum_{y=1}^{q} x_{ij} r_{ijy} \leq B_{uplink} \tag{3}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{p_i} x_{ij} r_{ijy} \leq R_y, \; y = q+1, \ldots, k \tag{4}$$

$$\sum_{j=1}^{p_i} x_{ij} r_{ijy} \leq R_{iy}, \; y = 1, \ldots, k; \; i = 1, \ldots, n \tag{5}$$

$$\sum_{j=1}^{p_i} x_{ij} = 1, \; i = 1, \ldots, n \tag{6}$$

$$x_{ij} \in \{0, 1\}, \; i = 1, \ldots, n; \; j = 1, \ldots, p_i \tag{7}$$

The Q-MOF may implement different algorithms and heuristics for solving the formulated problem (e.g., those proposed in [9]). In our service scenario (see Section V), we use the branch and bound algorithm. After having found an optimal solution, we search for a number of suboptimal but feasible solutions used to specify the MDP.

With regards to scalability, it is clear that for a large number of users, running the matching and optimization procedures separately for each service session may be too time consuming and costly. Thus, in addition to offering different levels of support, a solution may be for the Q-MOF to foresee certain “classes” of end users (users with common capabilities/preferences) and conduct matching and/or optimization procedures offline.

V. Example service scenario

To illustrate framework applicability, we consider an example service scenario involving a Web-based prototype application called Virtual Automobile Gallery (VAG) hosted by a third party SIP Application Server (AS). VAG allows a user to navigate through a 3D virtual gallery and view images of different automobiles. Throughout the world are stands that a user can click to view audio/video streaming clips. The application is offered in different versions and can support different audio and video codecs.

Assuming the network architecture shown in Fig. 9, we introduce the Q-MOF as a SIP Application Server to be included along the service control signaling path. The actors involved in the E2E QoS negotiation process include: end users, a primary SP (responsible for coordinating the negotiation process), the IP access network operator (taking on the role of a NP), and the VAG provider (taking on the role of a 3rd party application provider).

Using the Data Specification Model, we present an object diagram representing an instance of the VAG Service Profile in Fig. 10 (only relevant parameters shown for simplicity). The profile specifies three different service

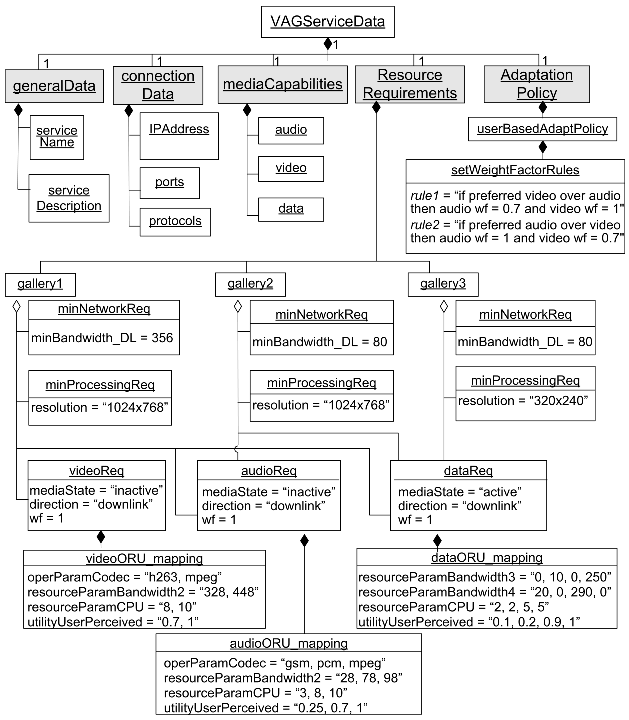

Figure 10. Object diagram for VAG Service Profile

versions: (1) Gallery 1 includes audio/video streaming and data (3D virtual scene) download; (2) Gallery 2 includes audio streaming and data download; and (3) Gallery 3 includes only data download. Consequently, the versions differ in network and processing requirements. For each media flow, resource requirements are expressed using an O-R-U mapping object.

During QoS (re)negotiation, the Matching Process is conducted by the Q-MOF to determine feasible parameters, followed by the Optimization Process to determine the ASP. We illustrate a case of invoking the Optimization Process when the negotiated service version is Gallery 1, and the optimization is invoked at a point in the session when both audio and video streaming are active. The problem is formulated as shown in Section IV-A. Resource vectors and corresponding utilities for streams are shown in Table I (based on the O-R-U mapping specified in the VAG service profile). The given operating points correspond to audio ($p_1 = 3$) and video streams ($p_2 = 2$). In this example we assume that cost is based on amount of traffic and assigned QoS class. We therefore calculate cost as *bandwidth_class_q* [bit/s] $\cdot$ *price_class_q* [monetary_unit/bit]. For the QoS class indicated as required for audio/video streaming, we assume a hypothetical price of 7 [monetary_unit/bit].

We specify a User Profile indicating that video is preferred over audio. The Service Profile adaptation policy is consulted which indicates that the audio WF w_1 should be set to 0.7 and the video WF w_2 to 1. The following constraints are determined based on the User Profile: maximum downlink bandwidth 620 kbps, maximum CPU 50%, and maximum cost 3700 [monetary_unit/s]. The result of the optimization is the ASP (Figure 11) which is

TABLE I

RESOURCE VECTORS AND CORRESPONDING UTILITIES FOR VAG

Resource vectors	Bandw. QoS class 2 [kbps] r_{ij1}	% total CPU r_{ij2}	Cost [monetary_unit/s] r_{ij3}	Utility $u_i(\mathbf{r}_{ij})$
$r_{11}(audio)$	28	3	196	0.25
$r_{12}(audio)$	78	8	546	0.70
$r_{13}(audio)$	98	10	686	1.00
$r_{21}(video)$	328	8	2296	0.70
$r_{22}(video)$	448	10	3134	1.00

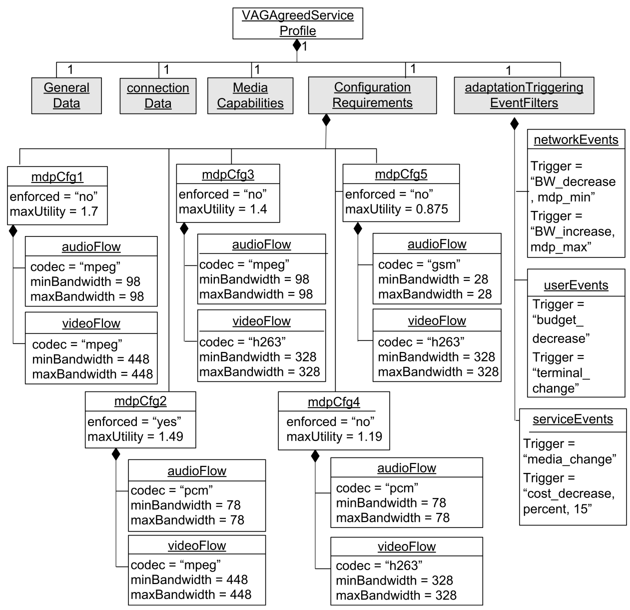

Figure 11. Object diagram for VAG Agreed Service Profile

signaled to indicate the agreed optimal service configuration and resource allocation, as well as an MDP. The MDP is composed of one optimal and four suboptimal service configurations. In addition to the MDP, a set of ATE filters are also specified based on the Adaptation Event Model, specifying various events leading to adaptation triggers.

For additional examples involving the Q-MOF in service-level QNA, including description of a laboratory prototype implementation, the reader is referred to our previous work [18] addressing applicability in a network based on the 3GPP IP Multimedia Subsystem (IMS) [2].

VI. DISCUSSION AND COMPARISON WITH RELATED WORK

The primary contribution of the proposed SELEQT framework is the comprehensive approach to solving and relating various aspects of the QNA problem for multimedia services by introducing business, data, adaptation, and negotiation models. The framework extends the NGN concepts and architectures proposed by ITU-T [1], 3GPP [2], and ETSI [3], by adding a Q-MOF in the service layer responsible for matching restrictive user, service, and network requirements, policies and constraints, with the goal of maximizing service utility. Two key steps comprising QNA have been introduced as a *Matching Process* (resulting in the specification of a FSP) and an *Optimization Process* (resulting in the specification of an

ASP). Furthermore, we present the idea of constructing MDPs and ATE filters for the purpose of efficient resource reservation and service adaptation.

The MDP considers user preferences and utility functions for all flows comprising a session when forming a degradation path, rather than calculating optimal adaptation for a single flow (as addressed in [10] and [17]). The ability to share the MDP information among the service and network layers might obviate the need for a trial-and-error approach to network resource reservation specified in existing NGN standards, hence leading to reduced signaling. In related work [9], optimal resource allocation among multiple sessions is formulated as a multi-choice multi-dimension 0-1 knapsack problem. We have applied these general concepts to our problem scope dealing with n parallel media flows belonging to a single session, and extended them by considering additional constraints and different QoS classes.

In terms of specifying various data relevant for negotiating QoS, the Data Specification Model may be considered as a reference model to be used by multiple actors involved in QNA for providing a common understanding of specified parameters and corresponding semantics. Existing session signaling standards (namely SIP/SDP) may be used to signal profiles as XML-based extensions. Mechanisms specified as part of the MPEG-21 multimedia framework standard [11] may also be integrated with session signaling and used in the context of service-level QNA, such as integration of the O-R-U mapping in the Service Profile. As compared to the E2ENP [6], used to signal alternative media configurations, the authors do not include specification of relative utility.

A number of significant international research projects deal with issues related to QNA, such as the DAIDALOS project [4] which proposes a Multimedia Service Provisioning Platform (MMSPP) acting as a mediator for terminal/application capability negotiation in heterogenous networks. While the MMSPP mainly leaves the determining of the final session QoS configuration to session end-points, our approach proposes that matching/optimization functionality be placed onto a network-based generic function (the Q-MOF), allowing for it to be reused by services. Furthermore, the ENTHRONE project [19] is investigating an integrated management solution supporting E2E QoS over heterogeneous networks and terminals, utilizing the MPEG-21 multimedia framework. The ENTHRONE solution does not support the notion of a FSP calculated prior to an ASP (serving as a basis for negotiation among entities) and signaling of an MDP.

Finally, considering different performance needs of various multimedia services (e.g., a mission critical telemedicine service vs. streaming music clips), it is clear that providing advanced QoS support will not always be necessary (or desirable) due to increased processing complexity and signaling overhead. The Q-MOF may be included along the signaling path when necessary, and designed to offer matching and optimization functionality at different levels, differing in complexity. For example, for certain services where dynamic adaptation will not be supported, the MDP will not be calculated.

VII. Current and Future Work

Our current work is addressing scalability related to implementing proposed support for E2E QNA in a real-world testbed, including the case of a large number of parallel sessions. Furthermore, we will investigate further aspects of the Business Model and impacts on negotiation (e.g., recognition, composition, trust elements etc.). With regards to Q-MOF, future research will explore the possibility for categorization of users and service requests in making domain-wide optimization decisions.

References

[1] ITU-T Y.2012 (2006) *Functional requirements and architecture of the NGN release 1.*

[2] 3GPP TS 23.228 Release 8 (2007) *IP Multimedia Subsystem (IMS); Stage 2*. Third Generation Partnership Project.

[3] ETSI TR 00001 V0.4.2 (2005) *TISPAN_NGN, Release 1: Release Definition.*

[4] IST DAIDALOS - Designing Advanced network Interfaces for the Delivery and Administration of Location Independent, Optimised personal Services (EU FP6), http://www.ist-daidalos.org/.

[5] Rosenberg, J. *et al.* (2002) SIP: Session Initiation Protocol. IETF RFC 3261.

[6] Guenkova-Luy, T. , Kassler, A.J. and Mandato, D. (2004) End-to-End Quality-of-Service Coordination for Mobile Multimedia Applications. *IEEE Journal on Selected Areas in Communication.*, vol. 22, no. 5, pp. 889-903.

[7] Handley, M. and Jacobson, V. (1998) SDP: Session Description Protocol. IETF RFC 2327.

[8] Araniti, G. *et al.* (2003) Adaptively Controlling the QoS of Multimedia Wireless Applications through 'User Profiling' Techniques. *IEEE Journal on Selected Areas in Commun.*, vol. 21, no. 10, pp. 1546-1556.

[9] Khan, S. (1998) Quality Adaptation in a Multisession Multimedia System: Model, Algorithms and Architecture. PhD Thesis, Dept. Of EE and Computing, University of Victoria.

[10] Mukherjee, D., Delfosse, E., Kim, J-G. and Wang, Y. (2005) Optimal Adaptation Decision-Taking for Terminal and Network Quality of Service. *IEEE Trans. on Multimedia*, vol. 7, Issue 3, pp. 454-462.

[11] ISO/IEC 21000-7 (2004) *Information Technology - Multimedia framework (MPEG-21) - Part 7: Digital Item Adaptation*. Moving Picture Experts Group.

[12] Skorin-Kapov, L. (2007) A Framework for Service-level End-to-End Quality of Service Negotiation and Adaptation. PhD Thesis, University of Zagreb, Croatia, 211 pp., December 2007.

[13] ITU-T E.860 (2002) *Framework of a service level agreement.*

[14] Hilt, V., Camarillo, G. and Rosenberg J. (2008) A User Agent Profile Data Set for Media Policy. IETF Work-in-Progress.

[15] Skorin-Kapov, L. and Matijasevic, M. (2007) A Data Specification Model for Multimedia QoS Negotiation. *Proc. of the ACM Third International Mobile Multimedia Communications Conference*, Nafpaktos, Greece, August 27-29, ACM.

[16] Skorin-Kapov, L. and Matijasevic M. (2005), "Dynamic QoS Negotiation and Adaptation for Networked Virtual Reality Services", Proc. of *Proceedings of IEEE WoWMoM*, Taormina, Italy, June 13-16, pp. 344-351. IEEE.

[17] Wang, Y., Kim, J-G., Chang, S-F. and Kim, H-M. (2007) Utility-Based Video Adaptation for Universal Multimedia Access (UMA) and Content-Based Utility Function Prediction for Real-Time Video Transcoding. *IEEE Transactions on Multimedia*, vol 9, Issue 2, pp. 213 - 220.

[18] Skorin-Kapov, L., Mosmondor, M., Dobrijevic, O. and Matijasevic, M. (2007) Application-level QoS Negotiation and Signaling for Advanced Multimedia Services in the IMS. *IEEE Communications Magazine*, vol. 45 no. 7, pp. 108-116.

[19] IST ENTHRONE - End-to-end QoS through Integrated Management of Content, Networks and Terminals (EU Framework Programme 6), http://www.ist-enthrone.org/.