

# Quality-assured and Sociality-enriched Multimedia Mobile Mashup

Hongguang Zhang, Zhenzhen Zhao, Shanmugalingam Sivasothy, Cuiting Huang, Noël Crespi

Wireless Networks and Multimedia Services Department

Institut Telecom, Telecom SudParis

9 Rue Charles Fourier, 91000, Evry, France

{hongguang.zhang, zhenzhen.zhao, shanmugalingam.sivasothy, cuiting.huang, noel.crespi}@it-sudparis.eu

**Abstract**—Mashups are getting more complex with the addition of rich-media and real-time services. The new research challenges will be how to guarantee the quality of the aggregated services and how to share them in a collaborative manner. This paper presents a metadata-based mashup framework in Next Generation Wireless Network (NGWN), which guarantees the quality and supports social interactions. In contrast to existing quality-assured approaches, the proposed mashup model addresses the quality management issue from a new perspective through defining the Quality of Service (QoS) metadata into two levels: fidelity (user perspective) and modality (application perspective). The quality is assured from quality-aware service selection and quality-adaptable service delivery. Furthermore, the mashup model is extended for users to annotate services collaboratively. The annotation occurs in two ways, social tagging (e.g., rating and comments), and QoS attributes (e.g., device type and access network etc.). In order to apply this network-independent metadata model into NGWN architecture, we further introduce a new entity named Multimedia Mashup Engine (MME) which enables seamlessly access to the services and Adaptation Decision Taking (ADT). Finally, our prototype system and the simulation results demonstrate the performance of the proposed work.

**Keywords**—Mashup, Next Generation Wireless Networking, Quality of Service, Sociality

## I. INTRODUCTION

The evolution of Web 2.0 has brought a significant impact on the Internet service provisioning by encouraging the contribution from end-user for contents and services creation. This phenomenon, termed User Generated Content (UGC) or User Generated Service (UGS), aim to enlarge user personalization through the “Do IT Yourself (DIY)” manner. Mashup, as a general term in the UGC/UGS domain, is an application that incorporates elements coming from more than one source into an integrated user experience [1]. Meanwhile, in Telecom there is an ongoing process of transformation and migration from so called legacy technology to an IP-based Next Generation Networking (NGN), or Next Generation Wireless Network (NGWN), which enables people to access multimedia anytime and anywhere. With the advantage of an All-over-IP network, the opportunity for integration and convergence is amplified, where the most prominent example is the Web-NGN convergence. Toward the convergence of

Web and NGN, mobile mashup is promising for the next generation user-driven multimedia delivery [2, 3].

With the proliferation of services available on the Internet and the emergence of user-centric technologies, millions of users are able to voluntarily participate in the development of their own interests and benefits by means of service composition [4]. The concept of composition is to create a new service by combining several existing elementary services. A number of composition mechanisms have been proposed, such as workflow technique and Artificial Intelligence (AI) Planning [5]. However, as most of the existing solutions are still professional developer inclined, the arduous development task always discourages users to contribute themselves to the service creation process. In this context, mashup, which is well known with its intrinsic advantages of easy and fast integration, is a promising choice for the user-driven service composition issues. Generally, the mashup mechanism is provided to combine non real-time Web services such as translation, search and map etc. by leveraging the programming Application Programming Interfaces (APIs). With the proliferation of mobile devices and wireless networks, real-time and resource-consuming multimedia services have been ubiquitous and all-pervasive. Thus, in this paper we consider mashup as user-driven multimedia aggregation. We argue that the user-driven multimedia delivery is more challenging than the provider-driven model. Firstly, to non-expert users it is desirable to have a mashup model which hides the backend complexity and simplifies the aggregation process. Moreover, the emerging mashups are getting more and more complicated when the rich-media and real-time services are aggregated. Nevertheless, the diverse terminals, heterogenous networks as well as various user requirements constrain multimedia mashup to low quality, especially in the mobile network environment. The third challenge is raised from the sociality point of view. Since the great success of social networking has shown that user experiences are enriched by sharing, aggregating, and tagging collaboratively, the social phenomena behind mashup are worth being explored.

Our paper presents a NGWN-based mashup framework, which is featured by an intermediate metadata model with the guarantee of quality and the support of sociality. The metadata-based framework brings the benefits in three aspects. Firstly, the human-readable metadata is the higher level description

language compared with programming APIs, which can hide the programming complexity from non-experts. Secondly, the scalable quality management can be enforced by Quality of Service (QoS) metadata. The concept of scalability in this paper means that the aggregated media can be tailored and adapted to diverse terminals and heterogeneous networks with the assured quality, which aims to provide the best user experiences across aggregated multimodal services. Thirdly, these metadata entities can be further enriched collaboratively by end-users through social annotation. In this paper, we propose to extend the CAM4Home metadata as our mashup model. CAM4Home is an ITEA2 project enabling a novel way of multimedia provisioning by bundling different types of content and service into bundles on the level of metadata [6]. In our solution, rich-media services including video, audio, image, and even text can be encapsulated as Collaborative Aggregated Multimedia (CAM) Objects, which can be then aggregated into CAM Bundles. We further propose to integrate MPEG-21 metadata within the CAM4Home model. We enforce QoS by two ways, quality-aware service selection at design-time, and quality-adaptable service delivery at run-time. The human-readable part of QoS metadata facilitates service selection firstly. Meanwhile it will enable adaptable delivery. Prominently, our system supports collaborative annotation. The annotation occurs in two ways, social tagging (e.g., rating and comments), and QoS tagging (e.g., device type and access network etc.). The former may facilitate service selection, while the latter will enhance QoS-aware mashup consumption.

The rest of the paper is organized as follows. Section II reviews the background and related works. In Section III, we describe a scenario and present the metadata-based model, in which we illustrate QoS management and social metadata. Section IV discusses the approach to apply the metadata model into the NGN-based service architecture. A prototype system and the performance evaluation are described in Section V. Section VI concludes the paper and presents some issues for future research.

## II. RELATED WORK

The past few years have witnessed the great successful of user-driven model, such as Wikipedia, Blog and YouTube, which is known as UGC. The next big user-driven hype will happen in the service arena, i.e. UGS. Considerable researches have been conducted on mashup and service composition, most of which utilize Web-based programming technologies (e.g., Web Service Description Language (WSDL) and Representational State Transfer (REST)) for the implementation. In order to facilitate the creation of mashup, some Web platforms have been proposed by different communities, among which Yahoo Pipes [7] and Microsoft Popfly [8] are well-known examples. Nevertheless, these platforms are far from being popularized for the ordinary users due to their complexity. It is desirable to have a mashup model which hides the backend complexity from user and simplifies the service creation interface, as well as satisfies the service creation variety requirements.

Unlike traditional data services, multimedia services face more challenges in the heterogeneous environments. A lot of research works have been conducted in this area. Z. Yu et al.

proposed a context-aware multimedia middleware which supports multimedia content filtering, recommendation, and adaptation according to changing context [9, 10]. The article [11] described an approach for context-aware and QoS-enabled learning content provisioning. L. Zhou et al. presented a context-aware middleware system in heterogeneous network environments, which facilitates diverse multimedia services by combining an adaptive service provisioning middleware framework with a context-aware multimedia middleware framework [12]. The scheduling and resource allocating issues were discussed for multimedia delivery over wireless network [13, 14]. However, these systems or solutions usually targeted one type of media. When more and more rich-media services are aggregated or composed, the quality issue is getting more challenging. In addition, the social phenomena between users are ignored by the past research works.

Typically, a mashup process can be divided into three steps: service selection, service aggregation and service execution. The quality issue is across these three steps, among which research efforts are firstly made to QoS-aware service selection. A composite service can be constructed and deployed by combining independently developed component services, each one may be offered by different providers with different non-functional QoS attributes. A random selection may not be optimal for its targeted execution environment and may incur inefficiencies and costs [15]. Therefore, a selection process is needed to identify which constituent services are to be used to construct a composite service that best meets the QoS requirements of its users. To formally define the QoS level required from the selected provider, the provider and the user may engage in negotiation process, which culminates in the creation of a Service Level Agreement (SLA). The management of QoS-based SLAs has become a very active area of research, including the QoS-aware service description, composition, and selection [16]. However, QoS-aware service selection is just the initial step to guarantee the quality. The other two steps may also bring a lot of impacts to the final quality. Most prominently, the context of service creation could be different to that of service execution, especially in the mobile environment, where the diverse terminals, heterogenous networks as well as various user requirements constrain the multimedia access to low quality. This problem is getting more and more complicated when the rich-media services are aggregated. As a result, a scalable model with QoS management is significantly important for mashups, especially for the mobile mashups in a highly dynamic service environment.

Since the mechanism of mashup is to combine data from different sources, it is desired to have an overall quality model across aggregated services. T.C. Thang et al. [17-19] have intensively studied the quality in multimedia delivery. They identified the quality from two aspects: perceptual quality and semantic quality. The former known as fidelity refers to a user's satisfaction, while the latter is the amount of information the user obtains from the content. The former is sometimes referred as Quality of Experience (QoE), while the latter is as Information Quality (IQ). In some cases, the perceptual quality of a media service is unacceptable or its semantic quality is much poorer compared with that of a substitute modality. A

possible solution for this problem is to convert the modality. For example, when the available bandwidth is too low to support the video streaming service for a football match, the text-based statistics service would be more appropriate than the adapted video with poor perceptual quality. This is a typical case of video-to-text modality adaptation. Apparently, the combination of fidelity and modality can enhance user experiences. Dynamic adaptation is seen as an important feature enabling terminals and applications to adapt to changes in access network, and available QoS due to mobility of users, devices or sessions [20]. The previous research works on multimedia adaptation are more concerned with the perceptual quality from the aspect of end user. However, the intensive studies in [17-19] state that the semantic quality should be considered in some cases. They argue that modality conversion could be a better choice than unrestricted adaptation on fidelity. The Overlapped Content Value (OCV) model is introduced in [17] to represent conceptually both quality and modality. Unfortunately, a quality model for mashup has never been mentioned in the literature. In this paper, we propose to apply both fidelity and modality into the quality of mashup. We argue that both perceptual quality and semantic quality need to be considered in order to provide quality-assure mashup.

Considering video as the most prominent media, we take video as the example for quality adaptation. There are some issues that cannot be ignored for video adaptation, such as complexity, flexibility and optimization. In this regard, Scalable Video Coding (SVC) has emerged as a promising video format. SVC is developed as an extension of H.264/MPEG-4 Advance Video Coding (AVC) [21]. SVC offers spatial, temporal and quality scalabilities at bit-stream level, which enables the easy adaptation of video by selecting a sub-set of the bit-stream. As a result, the SVC bit-streams can be easily truncated in spatial, temporal, and quality dimensions to meet various constraints of heterogeneous environments [19]. The three-dimensional scalability offers a great flexibility that enables customizing video streams for a wide range of terminals and networks. SVC can thus allow a very simple, fast and flexible adaptation to the heterogeneous networks and diverse terminals. M. Eberhard et al. have developed a SVC streaming test bed, which allow dynamic video adaptation [22]. It is desired to apply the advantages of SVC into mashup coping with the quality issue.

The ubiquitous multimedia results in the overwhelming multimedia services where it has become difficult to retrieve specific ones. Semantic metadata is a solution to the overwhelming resources. The lack of semantic metadata is becoming a barrier for the in-depth study and wide application. Recently, the great success of social networking has shown that user experiences are enriched by sharing, aggregating, and tagging collaboratively. Under this trend, folksonomy also known as social tagging or collaborative annotation draws more and more attention as a promising source of semantic metadata. Several works have been launched to exploit the knowledge of the mass in order to improve the composition process by considering either social networks or collaborative environments [23-25]. However, they only make use of sociality for service selection or recommendation. The sociality

across the process of mashup should be further explored, especially for the quality issue.

In this paper, we present a mashup framework as illustrated in Fig. 1. We enforce the quality by two ways, quality-aware service selection, and quality-adaptable service delivery. The proposed quality model considers both fidelity and modality to meet QoS requirements in the diverse terminals, heterogeneous networks as well as dynamic network conditions. We concentrate on both the user level by specifying user perceivable service parameter and the application level by adapting multimedia services according to the resource availability of terminal and network. Furthermore, we extend the mashup model allowing users to annotate the services collaboratively.

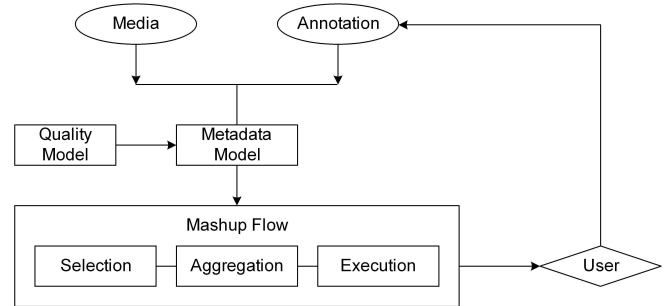


Figure 1. Mashup Model

### III. MASHUP MODEL

This section firstly describes the concept of metadata-based mashup model through an example scenario, followed by the illustration of the model decomposition. The mashup model is further decomposed into three essential parts: multimodal service aggregation, metadata-based QoS management and metadata-based social enrichment.

#### A. Concept of Mashup Model

Let's take "Sports Live Broadcasting" service as an example. The scenario is the last round of the football league where more than one team has the chance to win the champion. All teams start playing at the same time. Fans are watching the live TV broadcasting of their team. At the same time, they may also want to be updated on the information (e.g. goal, penalty, and red card etc.) of other simultaneous matches. We assume that there are two relevant services from different providers. The first one is an Internet Protocol TV (IPTV) program delivering a live football game. The IPTV service component can be configured by a set of offered alternative operating parameters (e.g., frame sizes, frame rates and bit rates etc.), by which IPTV can be adjusted dynamically according to user context. The second one is a real-time literal broadcasting service delivering statistics data synchronized to all football matches. A user composes the "Sports Live Broadcasting" mashup containing above two services. Before multimedia session, the quality model firstly selects the service version according to the static capabilities of terminals or networks. During session, this service element of IPTV can be adapted according to dynamic network condition or user preferences. Moreover, if the adapted IPTV service cannot provide the

expected user-perceived quality, a cross-modal adaptation from IPTV to Text may occur. Besides the quality adaptation, the fan can share the metadata-based mashup with friends like file sharing and annotate it by comment, rating as well as user-generated QoS parameters.

### B. CAM4Home Metadata

The essential part of mashup model is the multimodal service aggregation. In this paper, we use CAM4Home framework as the metadata model for multimodal service aggregation. The CAM4Home is an ITEA2 project implementing the concept of Collaborative Aggregated Multimedia (CAM) [6]. The concept of CAM refers to aggregation and composition of individual multimedia contents into a content bundle that may include references to content-based services and can be delivered as a semantically coherent set of content and related services over various communication channels. This project creates a metadata-enabled content delivery framework by bundling semantically coherent contents and services on the level of metadata. The CAM4Home metadata model supports the representation of a wide variety of multimedia content and service in CAM Element as well as its descriptive metadata. CAM Object is the integrated representation of CAM Element and CAM Element Metadata on the association rule “isMetadataOf”. CAM Bundles are the aggregation of two or more CAM Objects on the association rule “containsCAMObjectReference”. CAM Object and CAM Bundle can be uniquely identified by “camElementMetadataID” and “camBundleMetadataID”. Fig. 2 illustrates a conceptual view of CAM Bundle and CAM Object. Moreover, some complicated rules such as spatial and synchronization are also defined for enhanced aggregation.

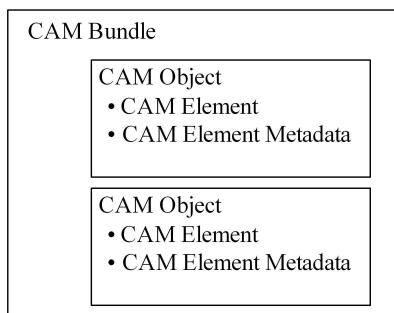


Figure 2. Conceptual view of CAM object and CAM bundle

The taxonomy of CAM Element has two subclasses, Multimedia Element and Service Element. The Multimedia Element is the container of a specific multimedia content, which is further divided into four types, document, image, audio and video. The Service Element is the container of a specific service. The physical content in CAM Element is referred by the attribute “EssenceFileIdentifier” which is a Universal Resource Locator (URL). The Service Element includes the other attribute “ServiceAccessMethod” indicating the methods used to access the service. With the instinctive of CAM, we use the metadata-based approach for the content and service delivery. The service capabilities are described by a CAM object containing Service Element and related metadata, while the converged service is described by a CAM bundle

containing several CAM objects of service capabilities. For instance, the attribute “EssenceFileIdentifier” can be used to indicate the Public Service Identity (PSI) of the service capability. And the other attribute “ServiceAccessMethod” indicates the SIP methods (e.g., INVITE) accessing the service. However, the described services are not limited to SIP-based. This model can be used to encapsulate any types of services. In this paper, the CAM4Home metadata model is adopted as the rich-media aggregation model. Fig. 3 shows an example for the aforementioned “Sports Live Broadcasting” service.

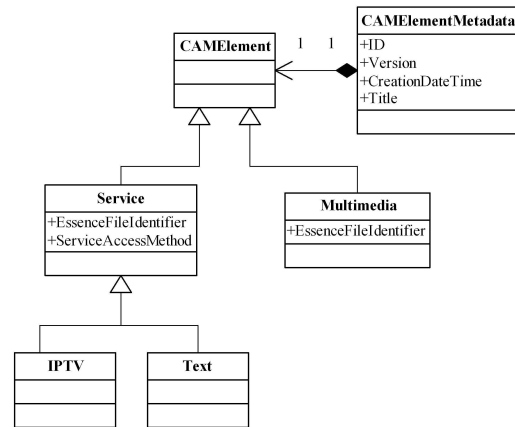


Figure 3. CAM4Home metadata example

### C. QoS Metadata

It is necessary to provide a quality-guaranteed and interoperable mashup delivery across various devices and heterogeneous networks as well as an optimized use of underlying delivery network bandwidth and QoS characteristics. Generally, it is a computing intensive process for adaptation decision-taking involved for choosing the right set of parameters that yield an adapted version. The computational efficiency of adaptation can be greatly enhanced if this process could be simplified, in particular by using metadata that conveys pre-computed relationships between feasible adaptation parameters and media characteristics obtained by selecting them [26]. Moreover, the development of an interoperable multimedia content adaptation framework has become a key issue for coping with this heterogeneity of multimedia content formats, networks and terminals.

Toward this purpose, MPEG-21 Digital Item Adaptation (DIA) specifying metadata for assisting adaptation has been finalized as part of the MPEG-21 Multimedia Framework [27]. MPEG-21 DIA aims to standardize various adaptation related metadata including those supporting decision-taking and the constraint specifications. MPEG-21 DIA specifies normative description tools in syntax and semantic to assist with the adaptation. The central tool is the Adaptation QoS (AQoS) representing the metadata supporting decision-taking. The aim of AQoS is to select optimal parameter settings that satisfy constraints imposed by a given external context while maximizing QoS. The adaptation constraints may be specified implicitly by a variety of Usage Environment Description (UED) tool describing user characteristics (e.g. user information, user preferences, and location), terminal

capabilities, network characteristics, and natural environment characteristics (e.g., location, time). The constraints can also be specified explicitly by Universal Constraints Description (UCD). Syntactically, the AQoS description consists of two main components: Module and Input Output Pin (IOPin). Module provides a means to select an output value given one or several input values. There are three types of modules, namely, Look-Up Table (LUT), Utility Function (UF), and Stack Function (SF). IOPin provides an identifier to these input and output values.

The mashup QoS management is proposed on two levels: fidelity and modality. The fidelity is to adapt one of the aggregated service component adjusting QoS parameters, i.e., multimedia adaptation with the perceptual quality from the perspective of end-user. The modality is to select the most appropriate modality among aggregated multimodal services components, i.e., modality conversion with the semantic quality from the application point of view. The overall quality model is illustrated in Fig. 4. We propose to integrate MPEG-21 DIA into CAM4Home model enabling QoS management. Originally in MPEG-21 DIA, the output values are utilized by Bitstream Syntax Description (BSD) for content-independent adaptation. However, in the proposed mashup model the adapted target is altered to CAM Bundle. Specifically, the AQoS is embedded in each CAM Object for quality adaptation as well as for modality adaptation. In this regard, for quality adaptation the output values (e.g. bit rate, frame rate, resolution) are utilized to yield an adapted version on a single service component.

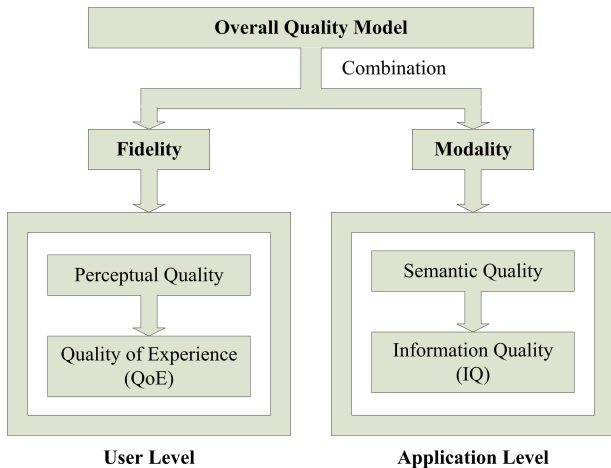


Figure 4. Mashup quality model

#### D. Social Metadata

Collaborative tagging describes the process by which many users add metadata in the form of keywords to shared content [28]. Social metadata is data generated by collaborative tagging, such as tags, ratings, and comments etc., added to content by individual users other than content creators. Examples can be found everywhere on the web, ratings and comments on YouTube, tagging in Digg. The social metadata can help users navigate to relevant contents even quicker because members can use them to provide context and relevant description to the content.

The proposed model takes advantage of social metadata to enrich the sociality of mashup from two aspects, service discovery and QoS management. Accordingly, users are allowed to annotate the services collaboratively in two ways: social tagging (e.g., rating and comments), and QoS attributes (e.g., device type and access network etc.). For example, Bob can tag a CAM entity that is relevant to him and choose the tags he believes best to describe the entity. The keywords Bob chooses help organize and categorize the service element in a way that is meaningful to him. Later, Bob or other members can use those tags to locate data using the meaningful keywords. In order to introduce ambiguous social tagging into structured metadata, the CAM4Home metadata framework defines some attributes of social metadata which include social tag, user comment, and user rating. As mentioned above as the second point, the QoS metadata can also be generated by users. For example, Bob can tag a CAM entity indicating the relevant service inside is not suitable for a mobile device with limited bandwidth. Usually, it is the service provider in the value chain of service delivery to take the responsibility on specifying these QoS parameters. However, it is cost-inefficient and time-consuming. The user-generated QoS metadata could be complementary to the provider-generated.

## IV. MOBILE MASHUP ARCHITECTURE

In this section, we firstly describe the mashup framework in detail. Then we propose the extension of session negotiation.

### A. NGN-based Mobile Mashup Framework

IP Multimedia Subsystem (IMS) has been widely recognized to be the service architecture for NGN/NGWN, offering multimedia services and enabling service convergence independent to the transport layer and the access layer. The IMS architecture is made up of two layers: the service layer and the control layer. The service layer comprises a set of Application Servers (ASs) that host and execute multimedia services. Session signaling and media handling are performed in the control layer. The key IMS entity in this layer is the Call Session Control Function (CSCF) which is a SIP server responsible for session control. There are three kinds of CSCF, among which Serving CSCF (S-CSCF) is the core for session controlling and service invocation. Home Subscriber Server (HSS) is the central database storing the subscriber's profile. Regarding the media delivery, the key component is Media Resource Function (MRF) that can be seen as media server for content delivery.

The IMS-based mashup framework firstly supports the combined delivery of multimodal services based on CAM4Home model. Further, the QoS management enforced by MPEG-21 DIA metadata is applied into IMS service architecture. Especially, the cross-modal adaptation is implemented as service switching among aggregated services. AS also interacts with MRF in order to ensure the adaptive delivery of media. Fig. 5 illustrates the conceptual mashup framework in IMS. The essential component in the proposed mashup platform is Multimedia Mashup Engine (MME) shown in Fig. 5. MME provides the controlled network environment between the mashup clients and the service repository. MME enables easy and seamless access to the service repository, and

supports the delivery of quality-assured experiences, across various devices, heterogeneous access networks and multiple service models (e.g., Web-based, Telco-based). Aforementioned, mashup is a user-driven model for service delivery. Therefore, MME is firstly proposed as a generic component of Service Deliver Platform, responsible for service-related functionalities, such as service registration and service discovery etc. Services represented as CAM metadata entities (e.g. object or bundle) are registered in MME. To end users, the rich semantic information may facilitate service composition and service discovery. The service repository holds both service objects and service bundles. To be noted that the service repository can be in MME or in an external database alternatively. For instance, the CAM4Home project provides a web service platform for metadata generating, storing and searching. In this case, MME needs to access the external platform through Web service interfaces.

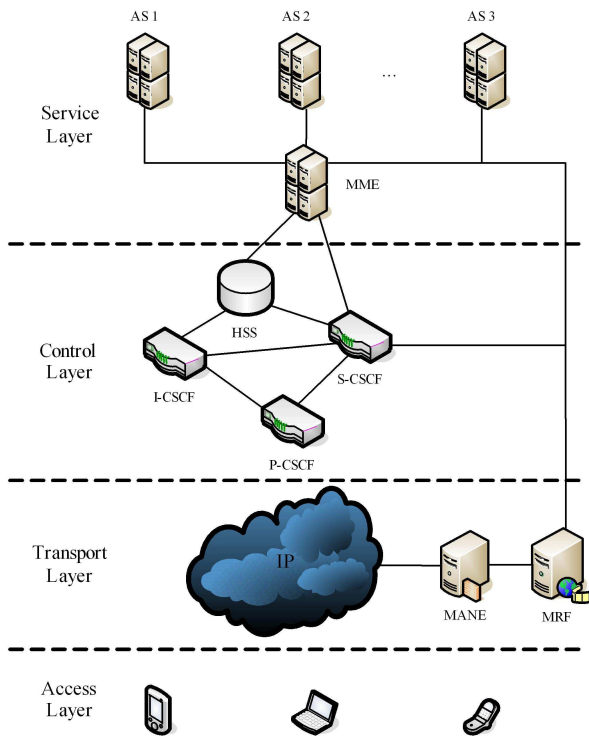


Figure 5. Conceptual mobile mashup framework

Besides above functionalities, the vital role of MME is service routing. MME provides the address resolution decision-making on ASs. As shown in Fig. 5, MME is located between S-CSCF and AS. For the consideration of scalability and extensibility, we collocate MME in a SIP AS behaving as Back-to-Back User Agent (B2BUA). One hand, MME is configured to connect with IMS. On the other hand, MME interfaces with SIP ASs which host those aggregated service elements. In order to enable quality-assured mashup, we extend MME mainly from three aspects: Adaptation-Decision Taking Engine (ADTE), UED collecting and social metadata interface. ADTE either selects appropriate content modalities among the aggregated service components or to choose adaptation

parameters for a specific media service. Additionally, MME needs to collect UED as inputs of ADTE. For modality selection, MME can act on the incoming requests and route them to AS according to the outputs of ADTE. Thanks to MPEG-21 QoS management, it is more intelligent compared with the routing criterion in [29] where it is based on the user requested service element. Secondly, MME supports the social metadata interface, through which end-users may enrich the original CAM metadata collaboratively.

For quality adaptation, we hereafter take video as the target considering video is the most challenging media type. We introduce the Media Aware Network Element (MANE), as shown in Fig. 5. The concept of MANE is defined as network element, such as a middlebox or application layer gateway that is capable of adapt video in real-time according to the configuring parameters. It is desirable to control the data rate without extensive processing of the incoming data, e.g., by simply dropping packets. Due to the requirement to handle large amounts of data, MANEs have to identify removable packets as quickly as possible. In our solution, the objective of MANEs is to manipulate the forwarded bit stream of SVC according to the network conditions or terminal capabilities. The target configurations of video that can be generated include bit rate, resolution and frame rate that in fact come as the outputs of ADTE.

### B. Session Negotiation Extension

The scalability we describe in this paper relies on the information exchange between client and server, which includes both static capabilities (e.g. terminal or network) and dynamic conditions (e.g. network or user preference). It allows participants to inform each other and negotiate about the QoS characteristics of the media components prior to session establishment. SIP together with Session Description Protocol (SDP) is used in IMS as the multimedia session negotiation protocol. However, the ability is very limited for SDP to indicate user environment information such as terminal capabilities and network characteristics. The User Agent Profile (UAProf) [30] is commonly used to specify user terminal and access network constraints. It is also not enough, because UAProf contains only static capabilities. Although RFC 3840 [31] specifies mechanisms by which a SIP user agent can convey its capabilities and characteristics to other user agents, it is not compatible with MPEG-21 based ADTE. It is important to reach interoperability between IETF approaches for multimedia session management and the MPEG-21 efforts for metadata-driven adaptation, in order to enable personalized multimedia delivery [32]. In our model, UCD and UED serve as the input of ADTE. These input values are in the format of XML document with a known schema. UCD includes the constraints imposed by service providers. We can assume that UCD is available for ADTE. However, UED should be collected for dynamic multimedia session in real time since it is the constraint imposed by external user environment. Therefore, there should be a way to query and monitor UED, particularly terminal capabilities and network characteristics.

In order to collect UED, we propose to extend the Offer/Answer mechanism. According to [33], SDP negotiation may occur in two ways, which are referred to as “Offer/Answer” and “Offer/Counter-Offer/Answer”. In the first way, the offerer offers an SDP, the answerer is only allowed to reject or restrict the offer. In the latter way, the answerer makes a “Counter-Offer” with additional elements or capabilities not listed in the original SDP offer. We slightly modify the latter way to put querying information in the “Counter-Offer”. DIA defines a list of normative semantic references by means of a classification scheme [34], which includes normative terms for the network bandwidth, the horizontal and vertical resolution of a display, and so on. For instance, the termID “6.6.5.3” describes the average available bandwidth in Network Condition. Table I show some examples of the semantic references. To indicate these normative terms in SDP, we define a new attribute/value pair as shown in Table II. “Offer” and “Answer” are distinguished by “recvonly” and “sendonly” respectively. The value in “Offer” means the threshold set by offerer, which is optional. The value in “Answer” is mandatory as return. In the adaptation framework, MME extracts the semantic inputs of AQoS and format them into SDP formats. During the Offer/Answer session negotiation procedure, the requested parameters are sent to UE in SDP. We assume that there is a module in User Equipment (UE) responsible for providing answers and monitoring dynamic conditions if necessary (e.g. presented by [35]). Accordingly, the answering values are also conveyed in SDP sending back to MME activating adaptation.

TABLE I. EXAMPLES OF SEMANTIC TERMID IN DIA

termID	Semantic References
6.5.9.1	The horizontal resolution of Display Capability
6.5.9.2	The vertical resolution of Display Capability
6.6.4.1	The max capacity of Network Capability
6.6.4.2	The minimum guaranteed bandwidth of Network Capability
6.6.5.3	The average available bandwidth in Network Condition

TABLE II. SDP EXTENSION

Method	Syntax
Offer (for query)	q= *(termID) a= *(recvonly:<value>)
Answer (as reply)	q= *(termID) a= *(sendonly:<value>)

The proposed adaptation process is divided into three phrases: session initiation, session monitoring and session adaptation. In the session initiation phrase, the party who invokes the service offers the default parameters in SDP by a SIP signaling message, normally SIP INVITE. Besides those well known parameters as answer, MME extracts input parameters in AQoS and offers them again as request. Some input parameters can be answered immediately such as terminal capabilities and network capabilities, which is enough for modality selection. However, some of them need

to be monitored in real-time, for example network conditions. In case that any parameter varies out of the threshold set by AQoS, a SIP UPDATE with the specific SDP is feedback to MME. Once ADTE in MME receives the inputs and makes a decision, the adaptation starts with session renegotiation. In case of quality adaptation, MME commands the MANE with the new parameters.

## V. PROTOTYPE AND EVALUATION

To verify the proposed approach, we develop a prototype system to demonstrate the scenario mentioned in Section III. The prototype system is the integration of several open source projects as illustrated in Fig. 6. On the server side, Open IMS Core [36] is deployed as IMS testbed. We make use of UCT Advanced IPTV [37] to provide IPTV service. MME and Text AS is set up by Mobicents SIP Servlet [38] and configured to connect with Open IMS Core. The client is simulated in the signaling plane and in the media plane separately.

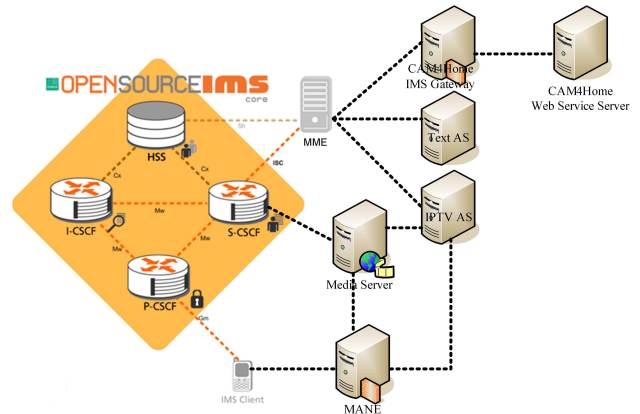


Figure 6. Prototype system

The CAM4Home metadata are central to the proposed mashup model. Aforementioned, the CAM4Home project provides a web service platform for metadata generating, storing and searching. In order to enable our client to access the service, we have deployed a gateway between IMS and CAM4Home. For metadata generating, a minimal set of data is required, such as title, description, essence file identifier etc. In our case, CAM objects with QoS metadata (e.g. IPTV and Text) are generated by service providers and deposited in the platform. End-users can search, aggregate, share, or annotate these multimedia resources through the gateway.

The system performance is analyzed in the signaling plane and in the media plane respectively. In the signaling plane, we emulate IMS signaling client by SIPp [39]. The prototype system demonstrates that the proposed SIP/SDP extension works compatibly with the standardized IMS platform. We observe that there are notably two kinds of latency: UED collecting and ADTE. The first one is more related to the characteristics of UED themselves. For instance, if the screen size is considered in UED, it could be retrieved immediately by UE. But in terms of available bandwidth, it depends on the time for sampling. Without considering UED, we further observe that ADTE-incurred delay is 100ms averagely. To some extent, this result confirms that the metadata-based

adaptation is efficient, because the pre-computation saves significant time over parameter selection.

The media plane is correlated with quality adaptation. We simulate three types of terminal with various resolutions: mobile phone, smart phone, and laptop. These terminals are assumed to be connected with three kinds of access networks, General Packet Radio service (GPRS), Universal Mobile Telecommunications System (UMTS) and Worldwide Interoperability for Microwave Access (WiMAX) respectively. The terminal settings are listed in Table III. The quality adaptation is simulated under the constraints of network bandwidth and terminal resolution. The SVC reference software JSVM 9.18 [40] is used as the video codec. The test sequence is ICE which is encoded with three spatial layers (QCIF, CIF, and 4CIF), five temporal layers (1.875, 3.75, 7.5, 15, and 30 fps) and two quality layers. The supported bitrates at various Spatial-Quality and Temporal are summarized in Table IV. Fig. 7 shows the average bitrates of adapted videos. Fig. 8 presents the output Peak Signal to Noise Ratio (PSNR) curves of adapted videos.

TABLE III. TERMINAL, ACCESS NETWORK AND SETTINGS

Terminal	Resolution	Access Network	Bandwidth
Mobile Phone	QCIF	GPRS	100 kbps
Smart Phone	CIF	UMTS	500 kbps
Laptop	4CIF	WiMAX	2000 kbps

TABLE IV. AVERAGE BITRATE

Bitrates [kbps]	1.875	3.75	7.5	15	30
QCIF	30.9 – 62.5	43.7 – 62.5	60.6 – 125.6	81.5 – 171.3	102.7 – 215.7
CIF	114.8 – 211.0	156.7 – 302.8	212.5 – 441.8	282.0 – 628.4	345.2 – 827.2
4CIF	353.0 – 700.7	497.8 – 1009.2	711.2 – 1503.4	1001.1 – 2211.0	1295.9 – 3046.0

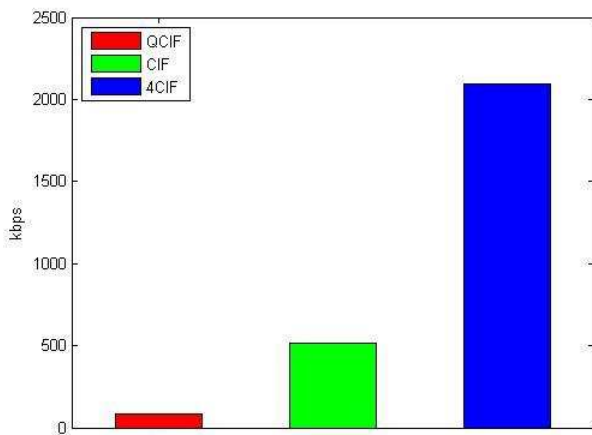


Figure 7. Output bitrate of adapted video

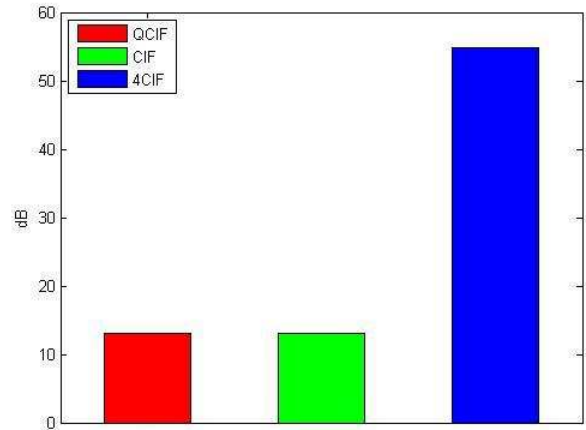


Figure 8. Output PSNR of adapted video

It can be seen that the average bitrates of adapted videos are consistent with the settings. And the adapted videos have different qualities, measured by means of PSNR. Obviously the bitrates co-relate with the values of PSNR. As we can see, SVC with the support ADTE is very suitable for quality-assured mashup. Considering this plane is more related to user experience, we plan to run usability tests in our future work.

## VI. CONCLUSION

This paper presented a metadata-based multimedia mashup framework in NGWN. It not only provided scalable QoS management but also enhanced the sociality of mashup. To achieve that, we proposed a flexible framework using the CAM4Home metadata model as a bundle of multimodal media. MPEG-21 DIA was further integrated into CAM4Home model to meet end-to-end QoS requirements. We addressed the issues in supporting QoS from two aspects, namely fidelity and modality, in order to tailor and adapt multimedia to the diverse terminals and the heterogeneous networks, as well as dynamic network conditions. The social annotations were used to enrich CAM4Home metadata collaboratively. Finally, a prototype system was developed on IMS architecture to validate the proposed model. With the use of rich metadata, context awareness and personalization could be challenging topics in the future.

## ACKNOWLEDGMENT

This research was supported by the Servery project and the CAM4Home project. We would like to thank all partners for their contributions. And thank Hui Wang and Mengke Hu for their simulation work.

## REFERENCES

- [1] D. Benslimane, S. Dustdar, A. Sheth, "Service Mashups: The Next Generation of Web Application," IEEE Internet Computing, vol. 12, Sept. 2008, pp 13-15.
- [2] A. Brodt, D. Nicklas, "The TELAR Mobile Mashup Platform for Nokia Internet Tablets," Proceedings of the 11<sup>th</sup> International Conference on Extending Database Technology, 2008.



- [3] B. Falchuk, K. Sinkar, S. Loeb, A. Dutta, "Mobile Contextual Mashup Service for IMS," The 2<sup>nd</sup> IEEE International Conference on Internet Multimedia Services Architecture and Applications, Dec, 2008.
- [4] X.Z. Liu, G. Huang, H. Mei, "A Community-centric Approach to Automated Service Composition," SCIENCE CHINA Information Sciences, vol. 53, no. 1, pp. 50-63, Jan. 2010.
- [5] J. Rao and X. Su, "A Survey of Automated Web Service Composition Methods," the First International Workshop on Semantic Web Services and Web Process Composition, July, 2004.
- [6] CAM4Home Official Website. Available: <http://www.cam4home-itea.org/>.
- [7] Yahoo Pipes, <http://pipes.yahoo.com/pipes/>
- [8] Microsoft Popfly, <http://www.popfly.com/com>
- [9] Z. Yu, X. Zhou, Z. Yu, D. Zhang, C. Chin, "An OSGi-Based Infrastructure of Context-Aware Multimedia Services," IEEE Communications Magazine, vol. 44, no. 10, pp. 136-142, Oct. 2006.
- [10] Z. Yu, X. Zhou, D. Zhang, C. Chin, X. Wang, and J. Men, "Supporting Context-Aware Media Recommendations for Smart Phones," IEEE Pervasive Computing, vol. 5, no. 3, pp. 68-75, Jul/Sep. 2006.
- [11] Z. Yu, Y. Nakamura, D. Zhang, S. Kajita, K. Mase, "Content Provisioning for Ubiquitous Learning," IEEE Pervasive Computing, vol. 7, no. 4, pp. 62-70, Oct. 2008.
- [12] L. Zhou, N. Xiong, L. Shu, A. Vasilakos, S. Yeo "Context-Aware Middleware for Multimedia Services in Heterogeneous Networks," IEEE Intelligent Systems, vol. 25, no. 2. pp. 40-47, Mar./Apr. 2010.
- [13] L. Zhou, X. Wang, W. Tu, G. Mutean, and B. Geller, "Distributed Scheduling Scheme for Video Streaming over Multi-Channel Multi-Radio Multi-Hop Wireless Networks," IEEE Journal on Selected Areas in Communications, vol. 28, no. 3, pp. 409-419, Apr. 2010.
- [14] L. Zhou, B. Geller, B. Zheng, A. Wei, and J. Cui, "Distributed Resource Allocation for Multi-Source Multi-Description Multi-Path Video Streaming over Wireless Networks," IEEE Transactions on Broadcasting, vol. 55, no. 4, pp. 731-741, Dec. 2009.
- [15] Q. Wu, A. Lyengar, R. Subramanian, I. Rouvellou, I. Silva-Lepe, T. Mikalsen, "Combining Quality of Service and Social Information for Ranking Services," The 7<sup>th</sup> International Joint Conference on Service-Oriented Computing, 2009.
- [16] V. Cardellini, E. Casalicchio, V. Grassi, F. Lo Presti, "Scalable Service Selection for Web Service Composition Supporting Differentiated QoS Classes," Technical Report, <http://www.ce.uniroma2.it/publications/RR-07.59.pdf>.
- [17] T.C. Thang, Y.J. Jung, Y.M. Ro, "Modality Conversion for QoS Management in Universal Multimedia Access," IEE Proceeding Vision, Image Signal Process, vol. 152, no. 3, pp. 374-384, 2005.
- [18] T.C. Thang, Y.J. Jung, Y.M. Ro, "Semantic Quality for Content-aware Video Adaptation," IEEE International Workshop on Multimedia Signal Processing, 2005.
- [19] T.C. Thang, J.G. Kim, J.W. Kang, J.J. Yoo, "SVC Adaptation: Standard Tools and Supporting Methods," EURASIP Signal Process: Image Communication, vol. 24, pp. 214-228, 2009.
- [20] M. Eberhard, L. Celetto, C. Timmerer, E. Quacchio, "An Interoperable Multimedia Delivery Framework for Scalable Video Coding based on MPEG-21 Digital Item Adaptation," IEEE International Conference on Multimedia and Expo, 2008.
- [21] H. Schwarz, D. Marpe, T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 9, September 2007.
- [22] A. Szwabe, A. Schorr, F.J. Hauck, A.J. Kassler, "Dynamic Multimedia Stream Adaptation and Rate Control for Heterogeneous Networks," Journal of Zhejiang University SCIENCE A, vol. 7, pp. 63-69, January 2006.
- [23] D. Schall, H. Truong, S. Dustdar, "Unifying Human and Software Services in Web-scale Collaborations, IEEE Internet Computing," IEEE Internet Computing, vol. 12, no. 3, 2008.
- [24] A. Maaradji, H. Hacid, J. Daigremont, N. Crespi, "Social Composer: A Social-Aware Mashup Creation Environment," ACM Conference on Computer Supported Cooperative Work, Feb. 2010.
- [25] M. Treiber, K. Kritikos, D. Schall, D. Plexousakis, S. Dustdar, "Modeling Context-aware and Socially-enriched Mashup," the 3<sup>rd</sup> International Workshop on Web APIs and Services Mashups, Oct. 2009.
- [26] D. Mukherjee, E. Delfosse, J.G. Kim, Y. Wang, "Optimal Adaptation Decision-Taking for Terminal and Network Quality-of-Service," IEEE Transactions on Multimedia, vol. 7, no. 3, pp. 454-462, June 2005.
- [27] Information Technology – Multimedia Framework (MPEG-21) – Part 1: Vision, Technologies, and Strategy, 2002.
- [28] S. A. Golder, and B. A. Huberman, "Usage Patterns of Collaborative Tagging Systems," Journal of Information Science, vol. 32, pp. 198-208, 2006.
- [29] H. Zhang, H. Nguyen, N. Crespi, S. Sivasothy, T. A. Le, H. Wang, "A Novel Metadata-based Approach for Content and Service Combined Delivery over IMS," the 8<sup>th</sup> Conference on Communication Networks and Services Research, May 2010.
- [30] OMA-UAPROF, "User Agent Profiling Specification (UAPROF) 1.1", Open Mobile Alliance, 12 December 2002.
- [31] IETF RFC 3840, "Indicating User Agent Capabilities in the Session Initiation Protocol (SIP)," August 2004.
- [32] A. Kassler, T. Guenkova-Luy, A. Schorr, H. Schmidt, F. Hauck, I. Wolf, "Network-Based Content Adaptation of Streaming Media using MPEG-21 DIA and SDPng," The Seventh International Workshop on Image Analysis for Multimedia Interactive Services, 2006.
- [33] IETF RFC 3264, "An Offer/Answer Model with Session Description," June 2002.
- [34] A. Vetro, Ch. Timmerer, S. Devillers, "Information Technology – Multimedia Framework – Part 7: Digital Item Adaptation," ISO/IEC JTC 1/SC 29/WG11/N5933, October 2003.
- [35] T. Ozcelebi, I. Radovanovic, M. Chaudron, "Enhancing End-to-End QoS for Multimedia Streaming in IMS-Based Networks," The Second International conference on System and Networks Communications, August 2007.
- [36] T. Magedanz, D. Witaszek, K. Knuettel, "The IMS Playground@FOKUS an Open Testbed for Next Generation Network Multimedia Services," First International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities, 2005.
- [37] UCT Advanced IPTV, <http://uctimsclient.berlios.de/>.
- [38] Mobicents, <http://www.mobicents.org/>.
- [39] SIPp, <http://sipp.sourceforge.net>.
- [40] SVC Reference Software (JSVM Software), [http://ip.hhi.de/imagecom\\_G1/savce/downloads/SVC-Reference-Software.htm](http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm).