Autonomic Service Control In Next Generation Networks

Michael Kleis[#], Andreas Klenk[*], Benoit Radier°, Sanaa Elmoumouhi°, Georg Carle[*], and Mikael Salaun°

[*] *Fraunhofer FOKUS,*
Kaiserin-Augusta-Alle 31
10589 Berlin, Germany
michael.kleis@fokus.fraunhofer.de

[#] *Technische Universität München*
Boltzmannstraße 3
85748,Garching bei München, Germany
{klenk,carle}@net.in.tum.de

° *France Télécom R&D*
Avenue Pierre Marzin 2
22307 Lannion, France
{benoit.radier, sanaa.elmoumouhi, mikael.salaun}@orange-ft.com

Abstract

Current standardization efforts aim towards a unifying platform for fixed and mobile telecommunication services. The IP multimedia subsystem is advocated as the candidate for building Next Generation Networks (NGNs). However, the direction taken in standardization is towards a rather static architecture with centralized features. The downside is an expected increase in service management complexity and the need for highly specialized infrastructures. This paper presents an approach for improving service quality, scalability and reliability while facilitating service management towards self-managing Next Generation Networks. To approach this we utilize and combine functionality available in the network using a Peer-to-Peer based service composition mechanism. The construction of composed services is based on a service chain principle and incorporates information about available services, Quality of Service (QoS) and applicable Servie Level Agreements (SLAs).

1. Introduction

Fixed mobile convergence is a hot topic in telecommunications industry. An important building block for next generation converged networks is the IP Multimedia Subsystem (IMS) defined by 3GPP[1] and taken into account by TISPAN[2]. The IMS allows for different types of access technologies while allowing mobile usage as well as an easy service integration. The main approach in IMS standardization is to define functional components and interfaces. The technical realization of this architectural model is inherently centralized and usually demands for a careful administration and deployment. Even in the case that IMS components are very reliable, the failure of an IMS component can lead to service interruption. This fact combined with the increasing complexity of service provisioning can result in a high management and configuration overhead for future IMS based services and thus high costs. In addition it can be expected that the resources of IMS components have to be allocated for peak usage, and will most of the time be underutilized. Thus CAPEX and OPEX for new services can be high and as a consequence the IMS architecture may be in fact not as flexible as expected.

In contrast, the prospects of autonomic networking research are to allow the network to take care of itself and to resolve problems automatically. In fact, the success of Peer-to-Peer (P2P) technology for Voice-over-IP (e.g. Skype) has already proven the value of distributed self-organizing architectures for telephony. Thus the question arises: *If and how P2P and overlay*

[1] 3GPP: Third Generation Partnership Project

[2] TISPAN: Telecoms & Internet converged Services & Protocols for Advanced Networks

technology can be adopted for service platforms in NGNs?

In this paper we describe the approach followed by the research project *Situated Autonomous Service Control* (SASCO) to explore and develop a secure, overlay based platform for an autonomous service provisioning in NGNs. To address the above-named question we start with the premises from the viewpoint of a multimedia service provider. The core requirements of a solution covering multimedia processing as well as QoS aware transport and routing are low costs, low management and configuration complexity as well as scalability. Based on these requirements the core research challenge is the exploration of a self-* [10][12] system for service provisioning in future networks. In this paper self-* denotes self-configuring, self-organizing, self-managing and self-repairing. Extending the results published in [1] we concentrate on an overall picture containing the required core concepts. For a detailed discussion of technical aspects we refer to [13][14]. As one result, the aspired approach would change the way how subscriber, network operator and service provider interact in a beneficial way for all parties. In the past, two traditional business relationships with regard to service provisioning dominated:

1. A direct business relation between clients and service providers in networks based on the end-2-end principle [11] as e.g. the Internet. The network provider is offering essentially the same interface to the transport service to client and service provider.

2. A business relation between the client and an operator or the network provider as in networks based on the intelligent network principle. In such networks, new services have to be introduced either by the network provider itself or by a third party provider using a special interface (e.g. Parlay X defined by ETSI) offered by the network provider for this purpose.

We propose to combine the strengths of both principles with the aim of defining an architecture that can be the basis for future and autonomic networks. The resulting entity model is depicted in Figure 1. As a consequence the approach will allow:

- The service provider to concentrate on service/content provisioning and to abstract from transport or end user terminal related issues.
- The network provider to offer value added transport services as: media adaptation to client terminals and access technology, broadcast/multicast services, caching, as well as seamless services and connectivity for clients.

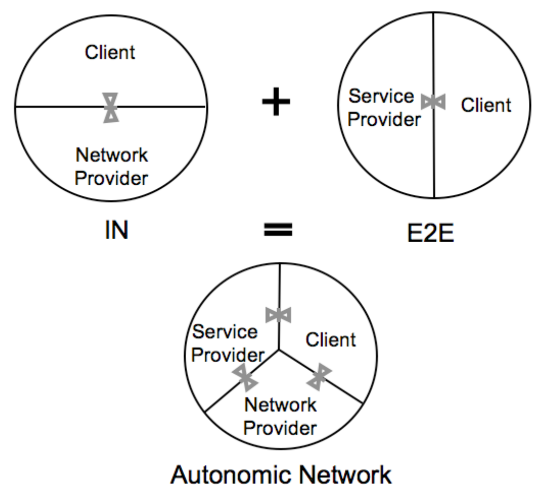

Figure 1 Entity Model

One of the main anticipated research challenges to realize this vision in the area of service composition is to resolve concrete service chains with a scalable distributed algorithm and to obey quality of service constraints imposed by the corresponding data transport and the services itself. In addition there is a need for explicit knowledge about the service chain to help the signalling between the partaking processing nodes and access control functions.

The paper is organized as follows: In Section 2 we present related work on autonomic overlay technology. Section 3 motivates the idea of autonomic service control. The following sections describe the decomposition and creation of service chains (Section 4), the DHT based control (Section 5), cooperative service provisioning (Section 6) and the integration of access control functionality (Section 7). We show how our work integrates withy existing IMS components in Section 8. We conclude in Section 9.

2. Related Work

In recent years one can observe a rising interest in overlay-related research. Since 2000, many classical network problems like QoS [24], Resilience [25], Multicast [26] or Security [23] have been addressed using the overlay approach. In addition, one branch of overlay-related research started to study the possibility of using an overlay concept for the flexible, on-demand composition of services.

One of the first projects in this direction was the Ninja Project [27]. The architecture developed by the project includes the notion of (logical and physical) service paths which have counterparts in many of the following proposals towards a overlay-based Service Composition as e.g. [28][29][30][31]. Further projects addressing service composition include SAHARA [34], SWORD [35] and the Ambient Networking Project [36]. The SAHARA project addressed trust and performance related aspects of service composition in case the component services are hosted by different providers. SWORD focused on the generation of service composition plans based on the requirements of the composed service with a strong focus on web service composition, while in Ambient Networks service composition was used for an on-demand processing and routing of media flows. Inside the overlay community the MONET group of Klara Nahrstedt at the University of Illinois Urbana-Campaign had a strong focus on network and QoS related questions [29][37][38][33][39][28][30][31]. Further contributions to this field can be found in [40][41] and [13]. In addition, from the perspective of active or programmable networks, routing problems closely related to the ones in overlay-based Service Composition have been addressed for example in [42] or [32]. Load balancing and stability issues for Service Composition have been discussed e.g. in [43].

In general, the overlay centric work towards a network and QoS aware composition of services can be classified into two main categories: 1. Centralized systems or systems which require global knowledge 2. Decentralized systems

Examples falling into the first category are [29][44][38][30][31][42][32]. As part of this category we consider approaches relying on a central point where service and QoS-related data is aggregated or schemes applying link state routing to address the Service Composition problem. In general, all schemes in this category require or assume permanent QoS measurements between all potential overlay nodes which results in a measurement overhead of $O(n^2)$ for n nodes. In addition it is necessary to either broadcast the measurement results and a description of offered services to the whole group, or to deliver them periodically to a central entity responsible for overlay setup.

This extensive measurement and dissemination overhead is required since in an overlay context the situation is different to the case of network layer link state protocols as e.g. OSPF. The reason for this is the fact that the nodes involved in overlays are usually end systems. For a given end system every other end system in the network is a potential overlay neighbor. Thus there is no notion of quasi-static network topology as in the case of layer three networking. Instead a Service Overlay topology is built up on demand and a link in the overlay in general corresponds to a path in the underlying network. As a consequence schemes in this category usually come with a large measurement overhead. In addition, a service composition problem involving multiple QoS metrics (as e.g. one additive and one concave metric) can already be considered as NP-Complete [45]. Therefore most schemes consider only one network-specific metric, or in case of two metrics use heuristics to reduce the complexity of the routing problems.

To summarize, the main drawback of this kind of schemes is the overhead introduced by the required periodic updates of QoS and service related information. As a consequence the size of addressable scenarios is limited with regard to number of nodes. On the other hand, since all required QoS information is collected proactively, the resulting service path calculation can be addressed directly after a request. Thus the schemes in this category in general have a short request response time.

Decentralized approaches as [40][33][39][44] and [13] address the service composition problem in a more reactive manner. As a consequence is it possible to address also large scale scenarios, since no central entity having global knowledge during the task of service overlay setup or for the discovery of valid processing chains is required. The approach proposed in this paper can be classified as decentralized. In fact we combine a DHT-based search for service components with on-demand QoS measurements. The actual measurements are used to verify that the QoS constraints of the requested composed service are not violated.

3. Service Overlays an Enabler for Autonomic Service Control

To establish overlay creation, maintenance and routing we start at the question: “How can the network provider take an active role in the provisioning of

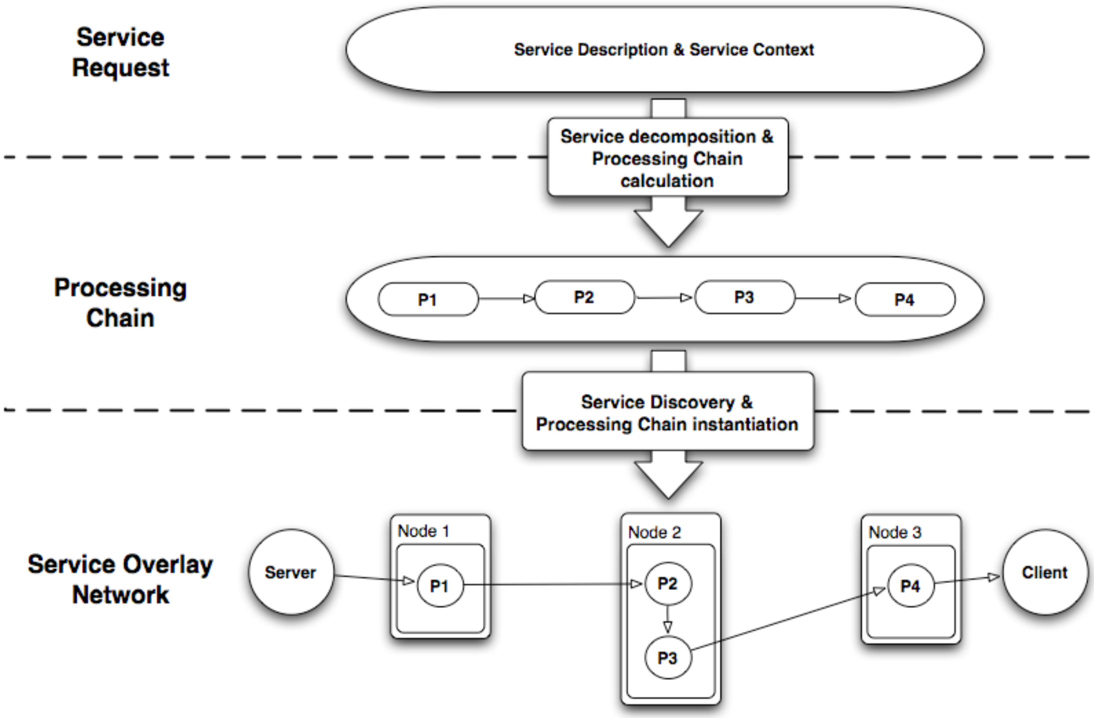

Figure 2 Main Steps for Service Overlay Creation

services in future network environments?" In fact by integrating the Network Provider into the process of service provisioning, QoS related problems can be addressed cooperatively by interaction between the Service Provider, the Network Provider and the Client. The reason for this is the fact that in such a case all entities involved in transport of data related to a service are also aware of the service itself. As a side effect, a Network Provider can be part of the service value chain e.g. providing value added transport to third party service providers as well as its clients. As an expected positive impact such an approach will allow

1. The Service Provider to concentrate on Service/Content and to abstract from transport or end user terminal related issues.
2. The Network Provider to offer value added transport services as: Media Adaptation to client terminal as well as access technology, Broad/Multicast services, Caching.
3. The Client to access services that are optimised for his/her end user terminal as well as access network technology.

In the paper we will have a strong focus on Service Overlays for Media/data transport and adaptation services. The main reason for this is the fact that Media Services have stringent QoS requirements i.e. demand QoS aware transport and processing.

4. Service Overlay Creation

In the remainder of this paper we assume that a Service Overlay consists out of an ordered sequence of processing modules interconnecting a service source and sink. The main focus will be on how principles from the area of Peer-to-Peer (P2P) networks can be used to realize an autonomous overlay creation by using service specific self-configuration of a distributed system of processing modules, clients and servers. Before we describe the proposed strategy, Figure 2 show the two main steps required for the creation of Service Overlays in such a scenario.
Given a service description the first requirement is a methodology to decompose a given service request into a set of distributable sub services. In general there are two main ways to address this:

- *Online decomposition:* i.e. decomposition of service at time of request
-
- *Offline decomposition:* i.e. decomposition during registration of a new service (i.e. in advance of the first request).

In this paper we propose to focus on *offline service decomposition* using a Service Level Agreement (SLA) principle. The SLA has to be established between a service provider and a third party provider (e.g. Network Provider) in advance of the first service request. The main reason for the SLA based approach is low complexity compared to the requirement of using service description languages to formulate respectively parse service request. A second advantage of the SLA approach is the fact that both parties can proactively optimize their server or network infrastructure in advance of the first service request based on the expected amount of service users.

After receiving a request for a decomposed service it is required to locate nodes possibly distributed inside and/or at the edge of the network hosting processing modules required for the instantiation of the requested service. To accomplish this *Service Discovery* task we maintain the information about available processing modules inside or at the edge of the network, using a Distributed Hash Table (DHT) (e.g. [17][18]19]). Based on the result of this service location step it is now required to interconnect the service source and sink through a sequence of processing modules (PM) using an Overlay Network principle in a way that the QoS constraints of the service are not violated and the costs of service provisioning are minimized. In the remainder of the paper a PM is formalized as a triple of the Form (I, P, O) where:

- "**I**" refers to the possible input formats the PM can read (i.e. Layer II/III/IV specific)
- "**P**" refers to the processing function provided by the PM
- "**O**" refers to the output format the PM produces (i.e. Layer II/III/IV specific)

Since it is assumed that neither the sink (e.g. Media Clients (MCs)) nor the source of the dataflow (e.g. Media Servers (MSs)) do any processing they are formalized using just an **(I,O)** notation. The MC, requesting content from a MS, can be served directly if and only if the input "**I**" of the client is *compatible* to the output "**O**" of the Server. In the case of non-compatibility, a PM has to be inserted between the MS and the MC to realize the data delivery using a pipelining principle. To denote compatibility the symbol "~" is used.

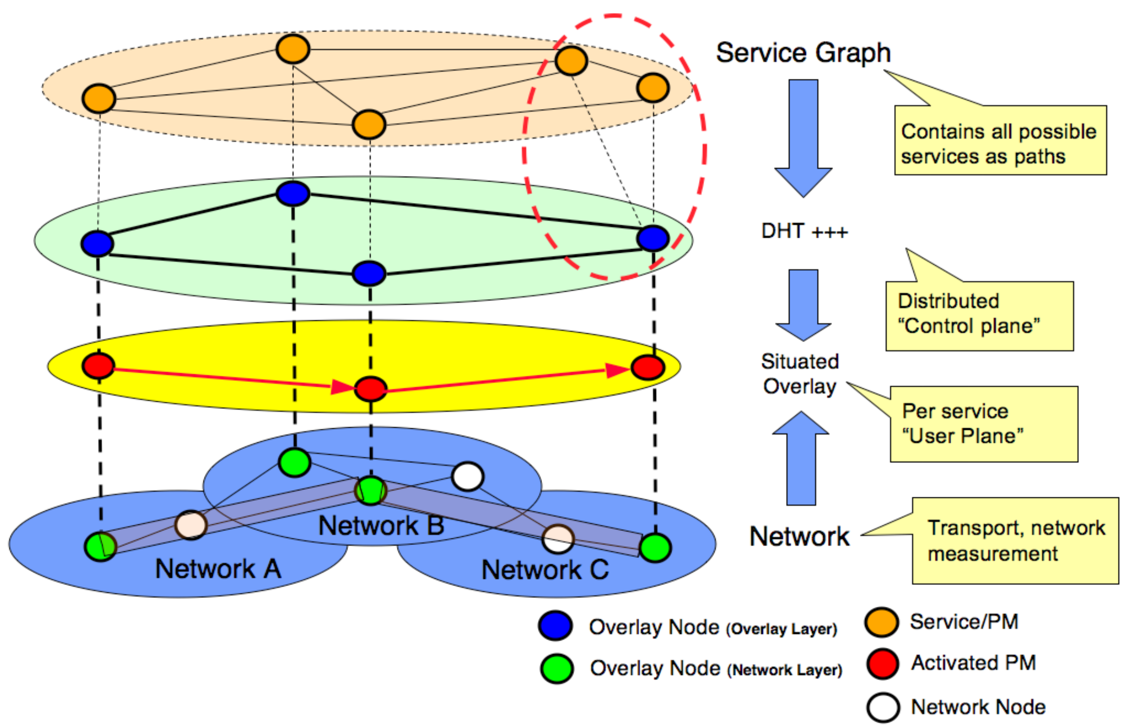

Figure 3 Different System Levels

4. A DHT based control plane for Service Overlay Creation

In Figure 3 we show the different system levels involved in the proposed service overlay creation process. Starting from the top, the set of all possible services to be realized can be modeled in a graph structure called *service graph*.

Definition (Service Graph): Let

$$V = \{PM_1, PM_2, \dots , PM_n\}$$

be a set of n processing modules. The Service Graph associated to V is defined as the graph SG(V, E) with

$$e = (PM_i, PM_j) \in E\text{:™ } PM_i < PM_j.$$

The property of a systems service graph we need in this paper is the fact that: *Every composed service that can be realized by a system corresponds to a path in its service graph*.

Using a problem specific indexing mechanism, the service graph structure is mapped into an address space of a Content Addressable Network (CAN) DHT [19]. This DHT will be extended towards a distributed control plane for the setup of service overlays.

The underlying idea of our approach is that in case every node that hosts processing functionality is also actively integrated into the search process, it is possible to build up the service overlay level while performing the search for its required processing functions. More concrete we are addressing Service Overlays creation based on a distributed CAN search principle combined with a hop-by-hop QoS constraint verification and propagation technique instead of extending classical routing algorithms as Dijkstra or Bellman-Ford (c.f. [20][21]). Using a DHT as the distributed control plane for a search & verify based approach has the following promising properties:

- *The resulting system can be realized in a fully distributed fashion and inherits the self-∗ properties of DHTs. Further it can be realized with comparable low management state per*

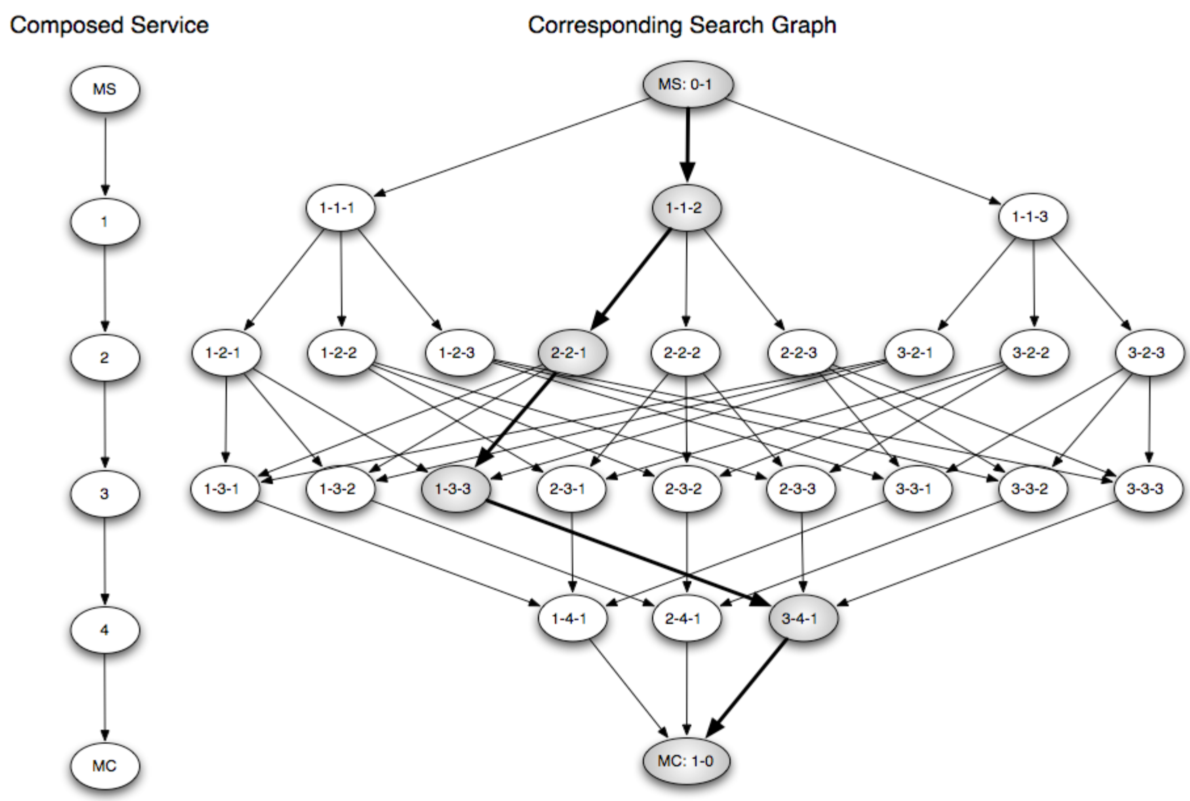

Figure 4 Search Graph

node e.g. O(log N), where N is the number of DHT nodes.

- *DHTs represent a well studied, resilient and fully decentralized domain for search based problems.*

5.1. Service Graph Embedding

As a prerequisite of the envisioned DHT based approach, it is required to specify how to embed a Service Graph structure into the corresponding DHT address space using a problem specific indexing scheme.

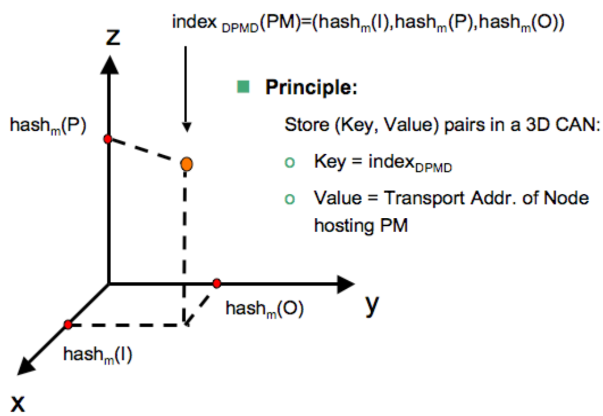

Figure 5 Indexing

To embed such a graph into a DHT address space we propose to focus on indexing functions that have the property that the above mentioned "~" relation is invariant with regards to the indexing process. We will illustrate this now by using a CAN DHT with address space $[0,t]\times[0,t]\times[0,t]\subset R^3$, for $t \in R$, t>0. After a new node n has joined successfully the CAN, the address where to store a pointer to the transport address where to find $PM_1 = (I_1 , P_1 , O_1)$ hosted by n, can be calculated as the coordinate

$$HASH_{CAN}(I_1,P_1,O_1) := (H(I_1), H(P_1), H(O_1))$$

for a hash function H having its values in [0,t] (c.f. Figure). The relation $\lesssim$ is invariant with regard to $HASH_{CAN}$ since:

$$PM_i \lesssim PM_j ™ HASH_{CAN}(PM_i) \lesssim HASH_{CAN}(PM_j)$$

In case each new PM made available to the system is registering itself at a CAN using the described $HASH_{CAN}$ function, we can find a PM_2 offering the processing function P_2 while being compatible to PM_1 by forwarded the search to all nodes in the CAN address sub space $(H(O_1),H(P_2), *)$ where "*" denotes any possible value.

5.2. Search & Verify

As stated before, every service that can be realized by a provider is corresponding to a path in its service graph. In case we want to realize a requested service it is required to find a corresponding path while taking the situation in the network into account. This task can be interpreted as a generalized Constraint Based Routing Problem (CBRP) including:

- a vector of QoS constraints related to the service
- a vector of constraints per Processing Step (e.g. delay introduced, costs etc.)
- a compatibility requirement between all the entities involved in service provisioning.

In its most general form such a CBRP can be formulated as:

Problem P1: Find an instance of the chain

$$(I_{MS},O_{MS}) \sim (I,P_1,*) \sim P_2 \sim \dots \sim P_{i-1} \sim (*,P_i,O) \sim (I_{MC},O_{MC})$$

while the constraints $C_{SID} = (C_1,...,C_n)$ are fulfilled.

We will approach P1 using a Search & Verify principle to distribute and parallelize the search for a solution of P1 between the nodes forming the CAN DHT layer. The basic principle of the proposed Search & Verify approach is shown using a simplified example with only one PM one additive metric as delay and one concave QoS metric as bottleneck bandwidth in Figure 6.

After the media source received the request for the service with $PC_{SID}=(P_1)$ it initiates a search for a processing module able to accomplish P_1. For each match a verify procedure is started measuring the values $H_1=(h^1_1,h^1_2)$ between the source and the PM, to verify the QoS parameters associated with the service, which have been specified via C_{SID}.
If $h^1_1 <= C_1$ and $h^1_2 <= C_2$, the corresponding PM is starting a new measurement task between itself and the

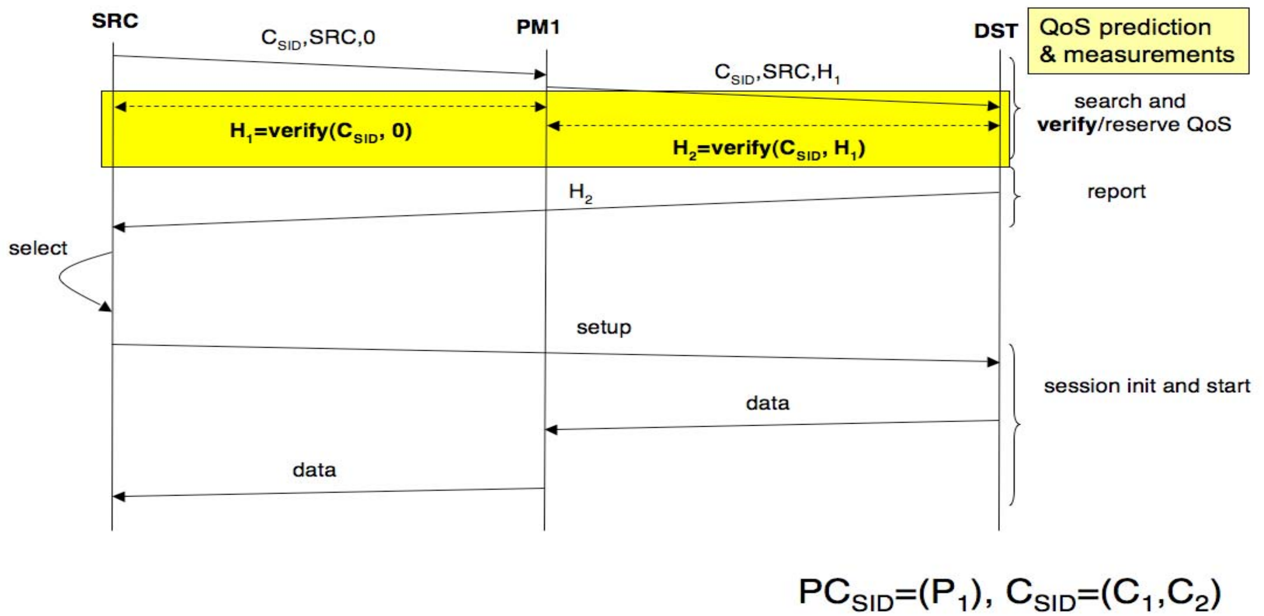

Figure 6 Search and Verify Approach

destination with result h^2_1, h^2_2. In case $h^1_1+h^2_1 <= C_1$ and $\min(h^1_2,h^2_2) <= C_2$ the destination is contacted and informed about the possible service chain found. In case not, all resources bound by the process are freed. In the simplified example, the client now reports all the possible service chains back to the source which selects the most adequate one based on QoS and cost values and initiates the data transfer.
In case of more complex services, the set of all possible processing chain candidates is in general defining a Directed Acyclic Graph (DAG) connecting the source and destination. We will call this DAG also the search graph associated with a service.
Figure 4 is showing an example search graph associated with the service P_{SID}=(1,2,3,4). For a more detailed study of the structure of search graphs and their relation to the complexity of the search and verify approach we refer to [13].

6. A Cooperative Service Provisioning Principle based on Service Overlays

The required interaction between Service Provider, Network Provider and Client for service provisioning based on the proposed system is divided into three phases, which are:

1. Service Registration
2. Service Request Processing
3. Service Delivery

In the following we will describe each phase in more detail.

6.1. Service Registration

During registration of a Service we establish a Service Level Agreement (SLA) regulating the QoS aware transport and processing requirements for a concrete service the Service Provider want to offer to a given Network Provider's customers. Since we focus on multimedia services we call this SLA from now on Multimedia Transport Service Agreement (MTSA). As stated before a Service Registration approach allows an offline decomposition of a service in advance of the first request.
The main steps to be performed during MTSA creation are based on our entity model and are illustrated in Figure 7 The minimum required set of information

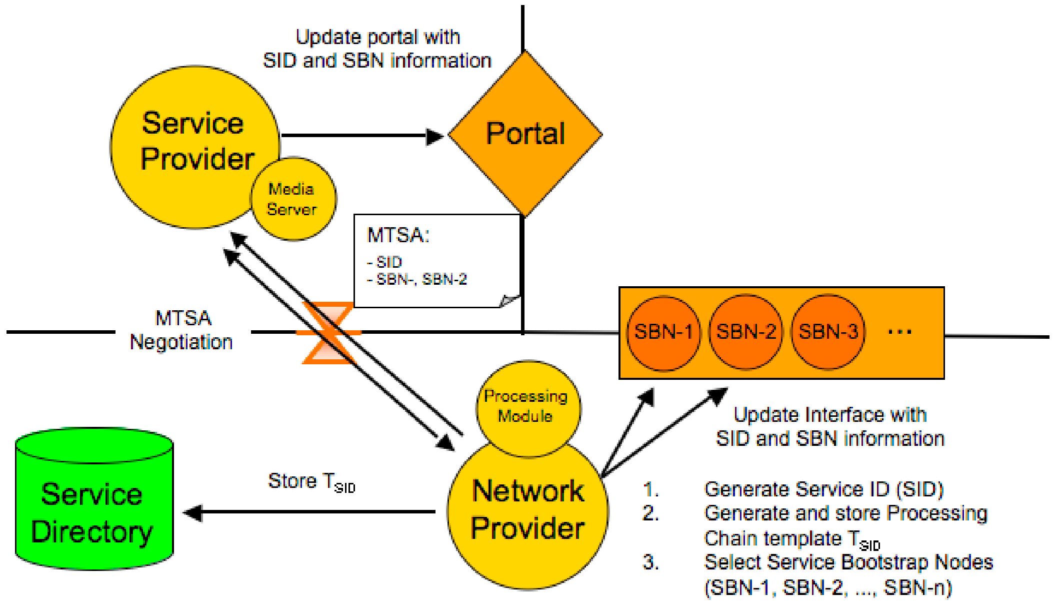

Figure 7 MTSA establishment

with regard to a new service to be negotiated during an MTSA agreement are:

- A unique Service ID (SID)
- A set of Service Sources (S_{SID}). E.g. a list of names or transport addresses of media servers hosting a special content
- A set of Service Bootstrap Nodes (SBNs) to initiate the Search & Verify process at the Network Provider.
- A set of required media/data processing steps (P_{SID}) to be performed by the Network Provider
- A set of constraints associated with the service (C_{SID}) as e.g. max. acceptable cost, max. acceptable loss or delay, bandwidth requirements etc.

We will use the term Processing Chain Template T_{SID} to denote the set of all information related to a service with id SID.

During a successful MTSA negotiation initiated by the Service Provider, the Network Provider is generating a SID and selects a set of Service Bootstrap Nodes (SBNs) to be used as entry points for accessing the new service. All this information is encapsulated into a T_{SID} which is stored at the responsible SBNs. After this step, the SID and the transport addresses of the responsible SBNs are sent to the Service Provider who can update e.g. a portal with information. From this point in time it is possible to access the service by connecting to a SBN responsible for the SID.
The decomposition of a service is done implicitly during the MTSA Negotiation process. To see this, we illustrate in Figure the main steps of an MTSA Negotiation. Essentially in Step 2 shown in the figure, the Network Provider is offering a set of service building blocks from witch the Service Provider is selecting a relevant subset at Step 3. After the selection the result is send to the Network Provider (Step 4). Therefore it is always guaranteed that the Network Provider can assign the list of processing steps P_{SID} to any negotiated Service SID.

6.2. Service Request Processing

Figure 9 shows the essential steps of a service request. From the portal of the Service Provider the client receives the required information to access a requested service i.e. the SID and the transport address of at least one SBN responsible for SID. With this information, the client can now send a service request to the SBN. As soon as the SBN has received the request, it performs a lookup operation for the corresponding T_{SID} extracting all information required to instantiate the service.

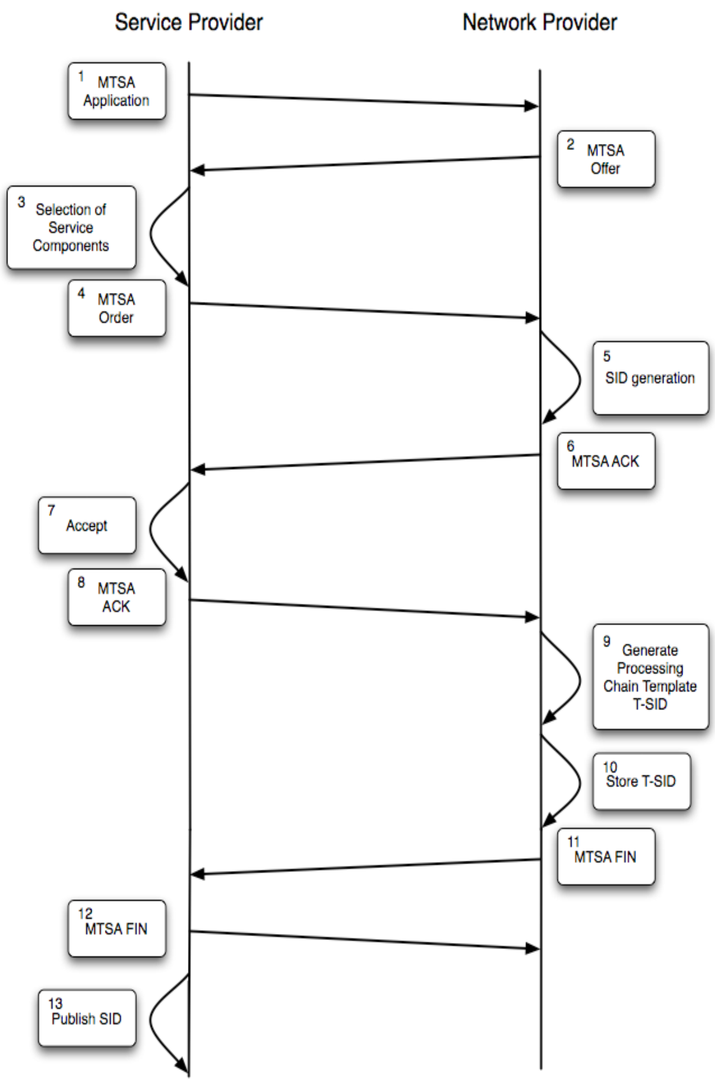

Figure 8 MTSA Negotiation

6.3. Service Bootstrap

At the point in time the Service Bootstrap is about to start the SBN has information about:

- **Service Specific**: Processing Chain $P_{SID}=(P_1,...,P_i)$, Constraints $C_{SID}=(C_1,...,C_n)$
- **Client Specific**: An acceptable input format I_{MC}
- **Server Specific**: At least one source with O_{MS}
- **Ingress PM Specific**: Required I and P_1
- **Egress PM Specific**: Required O and P_i

Based on this Information, the SBN can now initiate the creation of a Service Overlay network for SID. The SO will interconnect the Service Source, all required processing modules and the Client.

While all the former steps are in general based on the specification of an information model for service and network related data as well as the selection or design of suitable signaling protocols, this step is considered as the core problem to be addressed and was defined as P1.

7. Access Control in the Service Chain

One critical issue for a provider with regards to P2P based overlay technology is the fact that security and access control is often not an integral part of overlay networks [4]. A service platform without access control is incomplete. P2P research primarily perceived firewalls as an obstacle for the mutual connections between the participating overlay hosts [2][3] some overlays even disguise their traffic and tunnel through firewalls [16]. However, firewalls are successful security components that effectively guard services in the protected domain from unauthorized access. Packet filter firewalls are the predominant firewall architecture today. They base their decision on information from the headers of data packets, mainly at the network layer and the transport layer, but do not work well for hop by hop service chains. Traditional firewalls can only identify packets within the service chain that originate from the prior hop and are destined at the successor hop. They can neither identify the service source (the first service in the service chain) nor the service sink.

Firewalls loose their protective features if they cannot distinguish between legitimate and unauthorized overlay traffic. They will end up with a decision to either allow or to block all inbound overlay traffic.

The solution is to enhance firewalls for overlay networks to make them aware of service requestor, the

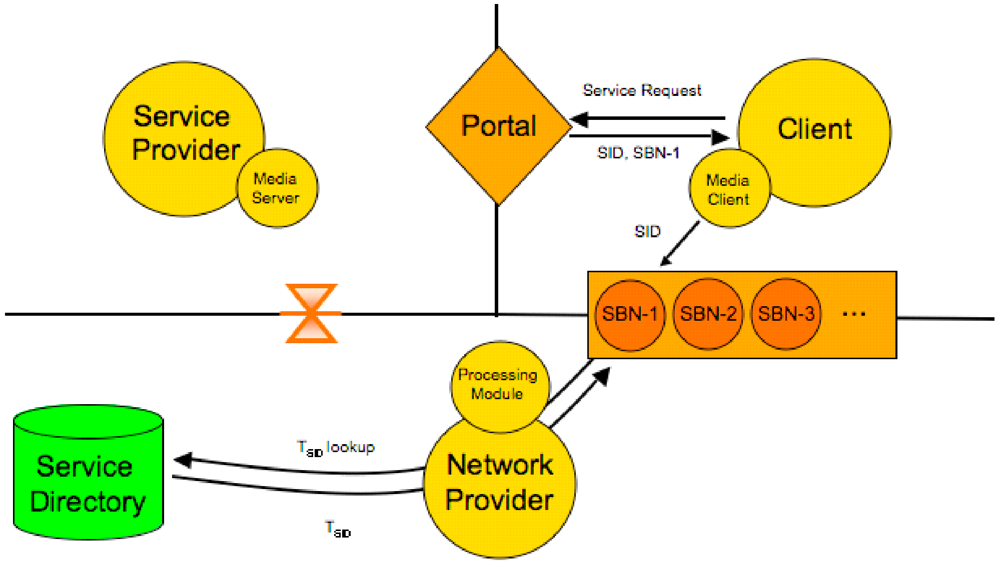

Figure 9 Service Request

processing modules and the destination of a service request. This new component must also be session aware to enforce fine grained access control decisions, for instance, a rule that a user is allowed to access a service only one time.

7.1 Overlay aware Access Control

Our approach is to extend the situated overlay by service chain aware access control functionality. There are two issues that require the integration with the overlay signalling.

1.) The authentication and authorization usually involves the cooperation with the service requester. Any service in a service chain might require authorization.

2.) Any access control function within the service chain must be aware of successor and predecessor and should also be able to distinguish and control traffic on a per session basis.

The integration of the overlay access control with the service chain instantiation solves these problems. The Situated Overlay discovers during the creation of a Service Overlay network for SID, that authorization is required at a number of services in the service chain. The Overlay supplies the access control functions with the next-hop and prior-hop address and describes the user session. Each MS, PM, and MC can either implement its own authorization policies, or rely on an authorization decision by another party (e.g. the network provider). Authentication deserves special attention in an overlay: Authentication is needed to verify subscriber identity.

7.2 Authentication

Authentication is responsible for asserting that the identity of the requester, as stated in the request, is correct. We give a brief overview how an overlay solution could integrate different NGN identity solutions for user authentication. NGN networks already have a sophisticated infrastructure for authentication in place (e.g. the Generic Authentication Architecture). The 3GPP defined with the TR 33.980 a coupling of the IMS with the Liberty Alliance - Federated Identity Management Framework. The coupling is done with the help of the Web Services Framework (ID-WSF) and the Federation Framework (ID-FF) for authentication. The 3GPP describes two options. The Network Application Function (NAF) is part of the IMS and can act as an Identity Provider (IdP) at the same time. The second option is to combine the NAF with the Authentication Service (AS) of the Web Services Framework. An overlay network can integrate these identity systems to obtain verified user identities.

8. P2P Principles for an Autonomic Service Control in the IMS

To be able to use P2P principles in the IMS context it is first required to address the question: What are

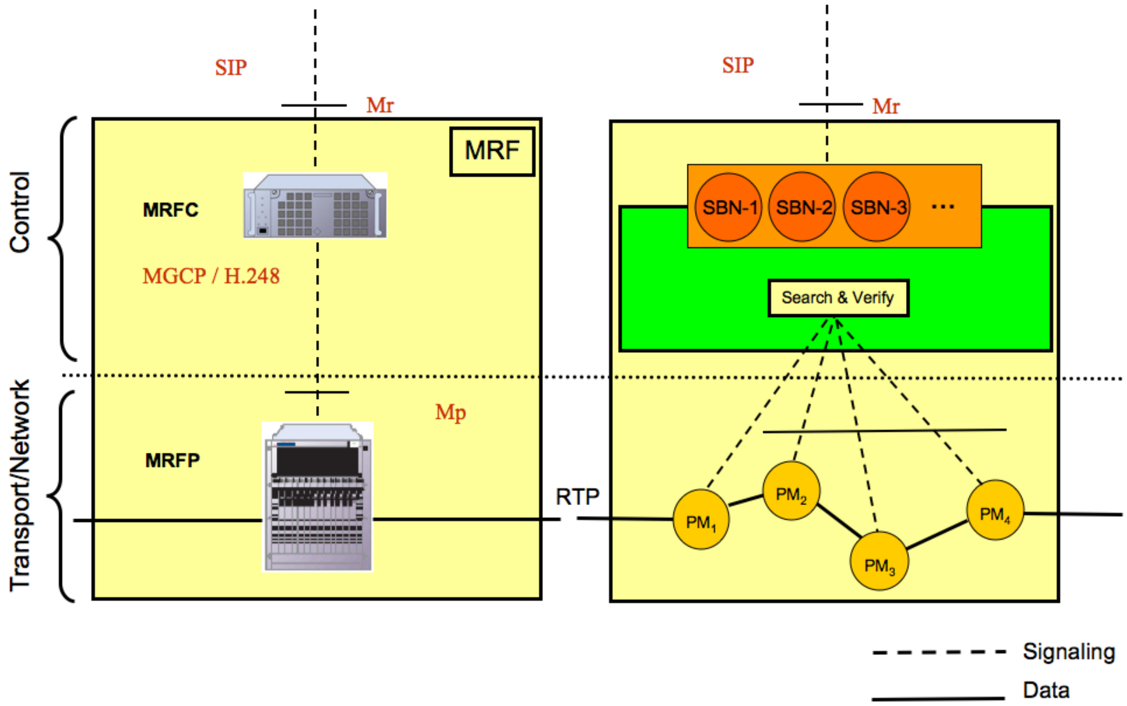

Figure 10 3GPP MRF

IMS functions that can be realised using P2P principles? As one candidate we identified the Media Resource Function (MRF) which is responsible for resource consuming tasks e.g. playing, transcoding and mixing of media streams. As an example from TISPAN, the Resource and Admission Control Subsystem (RCAS) realizes access and would profit from the proposed Situated Overlay approach by a better insight into service relationships for authorization decisions (refer to Section 7).

Another important issue is: What entities should run the corresponding P2P software? To address this, we categorise potential candidates into three main categories:

1. **Dedicated infrastructure nodes** are part of the network infrastructure and are particularly trustworthy and reliable.
2. **Home gateways and set top boxes** are fixed clients under partial control of the network provider that are already deployed in the order of several million devices in Europe.
3. **Mobile clients** have unsteady availability and would typically appear as service-originator or consumer in the situated overlay.

The IMS can be considered as a SIP based signaling overlay including AAA functions and support for multimedia calls and services. To enable the flexible introduction of new IMS based services a SIP Application server is included. To interface with an existing IMS it we propose the realization of an IMS Media Resource Function based on the described approach.

8.1. Realization of a distributed IMS Media Resource Function

The MRF represents the media related core of the IMS. It is composed out of the Media Resource Function Controller (MRFC) and the Media Resource Function Processor (MRFP). Since MRFC and MRFP are centralized entities, the same scalability and reliability problems as in case of a classical client/server approach for media delivery can be anticipated.

In this context the described approach can be used to realize a distributed MRF utilizing P2P principles. In addition it is possible to extend standard MRF functionalities as conferencing (mixing), recording and playback of static content (e.g. announcements) with additional services as

- Transcoding and/or downscaling of audio/video data,
- Traffic shaping to avoid congestion,
- Network based error correction techniques

by registering the corresponding processing functionality in form of PMs.

In Figure 10 we mapped the relevant control as well as network/transport related functionalities to a 3GPP MRF.

As illustrated in the Figure 10 the Mr interface can act as the interface between an IMS and the SBN component. As a prerequisite for this it is required that an instance of the Mr interface is integrated into SBN nodes. In the reminder of the section, we use a SIP based approach for MRF control described in [22] to provide an example how such an Mr interface can look like.

The MRF control protocol described in [22] utilizes SIP INVITE or SIP INFO messages as carrier for service request from a SIP Application Server (SIP AS) or the S-CSCF to the MRF. The actual service request and the required parameters are encoded into the SIP URI [22] contains specifications how to control *announcements*, *interactive voice response* (Prompt and Collect) and *conferences*.

After receiving and processing a service request, the MRF is sending back a return code to signal successful completion or failure of the service request. To illustrate this approach Figure 12 shows two possible SIP request URIs for an *announcement.*

```
sip:annc@ms.example.net; \
play-http://audio.example.net/announcement.g711

Or:

sip:annc@ms.examples.net; \
play=file://fileserver.example.net/bar/foo.wav
```

Figure 12 RFC 4240 Service indicator for Announcements

Announcements are media files played to a user. Following the description in the RFC “Announcements can be static media files, media files generated in real-time, media streams generated in real-time, multimedia objects, or combinations of the above.” In addition the

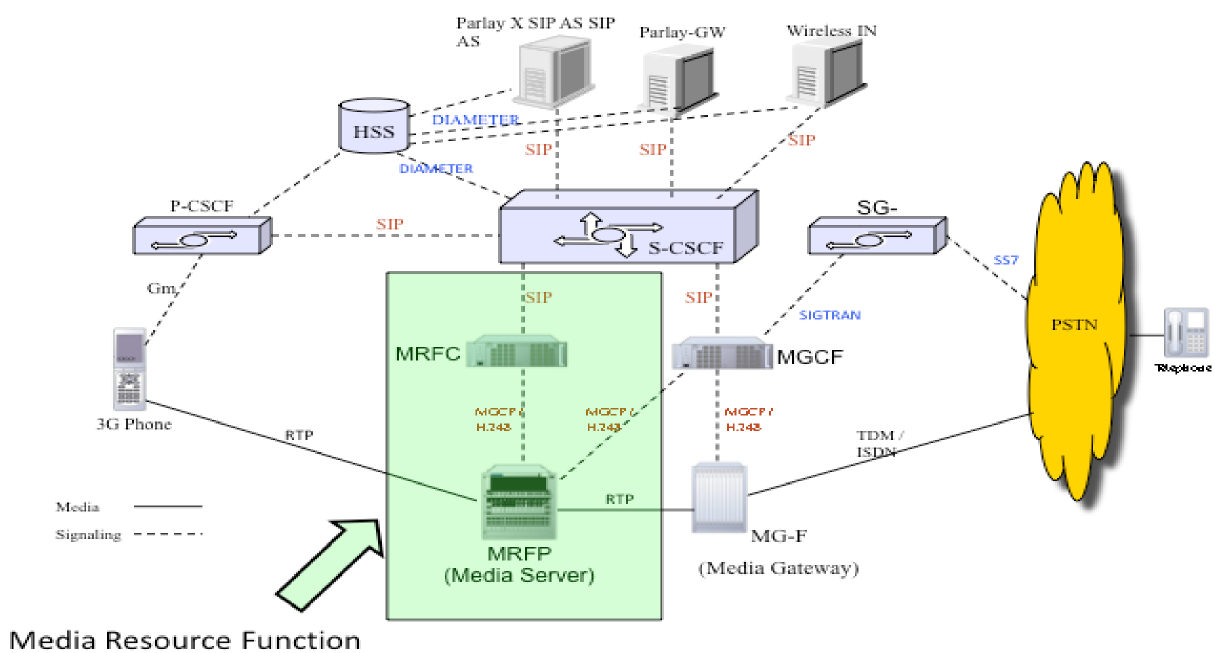

Figure 11 Media Resource Function in the IMS

authors state that "the mechanism described in this document has absolutely no impact to SIP devices other than media servers".

To adopt the announce command to start and stop Topic O services we need to transport:

- The ID of the Service requested (SID)
- Client IP and Port
- The Capabilities of the requesting Client (MC_I)

As a consequence a start message send from SIP-AS or S-CSCF can be of the form:

```
sip:annc@sbn.example.net; \
play=SASCO-START://SID/MC_TP:PORT/MC_T
```

The corresponding stop message can be defined as:

```
sip:annc@sbn.example.net; \
play=SASCO-STOP://SID
```

Based on this categorisation we propose to use a hybrid P2P IMS approach using a combination of infrastructure nodes and fixed clients for a P2P based IMS since these nodes can be assumed to have the necessary stability and connectivity to be integrated into a service provisioning process.

To illustrate this Figure 11 shows how a distributed IMS MRF can be realised using the situated overlay and the IMS Mr interface to link the situated overlay to the rest of the IMS.

9. Conclusion

We are at the brink of the realization of Next Generation Networks with IMS at the core. To be able to manage the expected increasing configuration complexity caused by the plethora of future services we are convinced that self-management capabilities are crucial for the success of IMS/TISPAN based technology. To reach this we propose to exploit the autonomic functionalities of peer-to-peer based overlay technology to form an Autonomic Service Control for Next Generation Networks. We identified suitable entities that could form the P2P overlay network and IMS functions that benefit from a realization as overlay service. We introduced a decentralized service composition mechanism that obeys quality of service parameters. We argued with an access control scenario how service chain knowledge facilitates the required signaling for authorization of service usage. Future work is to realize the Autonomic Service Control and evaluate its role for IMS/TISPAN architectures.

10. Acknowledgements

This research is performed in the context of the Situated Autonomic Service Control (SASCO) project, which is gratefully funded by the France Télécom R&D.

11. References

[1] Andreas Klenk, Michael Kleis, Benoit Radier, Sanaa Elmoumouhi, Georg Carle, and Mikael Salaun, "Towards Autonomic Service Control in Next Generation Networks", The Fourth International Conference on Autonomic and Autonomous Systems, ICAS 2008, March 16-21, 2008 - Gosier, Guadeloupe

[2] Baset, S.A. & Schulzrinne, H., 'An Analysis of the Skype Peer-to-Peer Internel Telephony Protocol', Arxiv preprint cs.NI/0412017 (2004).

[3] Ennis, D.; Anchan, D. & Pegah, M.,The front line battle against P2P, in 'SIGUCCS '04: Proceedings of the 32nd annual ACM SIGUCCS conference on User services', ACM Press, New York, NY, USA, pp. 101—106 (2004).

[4] Wallach, D.S.,A Survey of Peer-to-Peer Security Issues., in 'ISSS', pp. 42-57 (2002).

[5] D. Xu and K. Nahrstedt, "Finding service paths in a media service proxy network," Proceedings of the ACM/SPIE Conference on Multimedia Computing and Networking, (2002).

[6] K. N. Jingwen Jin, "Source-based QoS service routing in distributed service networks", Proceedings of IEEE International Conference on Communications 2004 (ICC2004), 2004.

[7] X. Gu, K. Nahrstedt, and B. Yu.'SpiderNet: An Integrated Peer-to-Peer Service Composition Framework',Proceedings of the thirteenth IEEE International Symposium on High-Performance Distributed Computing (HDPC-13) 2004

[8] P. Maymounkov and D. Mazieres, "Kademlia: A peer-to-peer information system based on the xor metric," in Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02), March 2002

[9] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, "A scalable content-addressable network," in Proceedings of the 2001 ACM SIGCOMM Conference, August 2001.

[10] Ganek, A. G. Corbi, T. A.: The dawning of the autonomic computing era. IBM Systems Journal. 42(1), 5—18 (2003)

[11] J. H. Saltzer, D. P. Reed, Anind Dey, Understanding and D. D.Clark. End-to-end arguments in system design. ACM Transactions on Computer Systems, pages 277-288, 1984.using Context, Personal and Ubiquitous Computing, 2001

[12] O. Babaoglu, M. Jelasity, A. Montresor, C. Fetzer, S. Leonardi, A. van Moorsel, and M. van Steen, Eds., Self-Star Properties in Complex Information Systems, ser. Lecture Notes in Computer Science, Hot Topics. Springer-Verlag, 2005, vol. 3460.

[13] Michael Kleis, Kai Büttner, Sanaa Elmoumouhi, Georg Carle, Mikael Salaun, "CSP, Cooperative Service Provisioning using Peer-to-Peer Principles" Proceedings of 2nd IEEE/IFIP International Workshop on Self-Organizing Systems (IWSoS), 2007

[14] Andreas Klenk, Frank Petri, Benoit Radier, Mikael Salaun, Georg Carle "Automated Trust Negotiation in Autonomic Environments" Proceedings of 2nd IEEE/IFIP International Workshop on Self-Organizing Systems (IWSoS), 2007

[15] "Telecommunications and Internet converged Services and Protocols for Advanced Networking" TISPAN; NGN Functional Architecture Release, ETSI ES 282 001 V1.1.1, 2005

[16] Androutsellis-Theotokis, S. & Spinellis, D., 'A survey of peer-to-peer content distribution technologies', ACM Comput. Surv. 36(4), 335—371 (2004).

[17] P. Maymounkov and D. Mazieres, "Kademlia: A peer-to-peer information system based on the xor metric," in Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02), March 2002

[18] I. Stoica, R. Morris, D. Karger, F. Kaashoek, and H. Balakrishnan, "Chord: A scalable Peer-To-Peer lookup service for internet applications," in Proceedings of the 2001 ACM SIGCOMM Conference, August 2001, pp. 149–160.

[19] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, "A scalable content-addressable network," in Proceedings of the 2001 ACM SIGCOMM Conference, August 2001.

[20] J. Jin, K. Nahrstedt. Source-Based QoS Service Routing in Distributed Service Networks. In Proc. IEEE International Conference on Communications 2004 (ICC2004), Paris, France, Jun. 2004.

[21] S Ralph Keller. Self-Configuring Services for Extensible Networks -- A Routing-Integrated Approach. PhD thesis, Computer Engineering and Networks Laboratory (TIK) ETH Zurich, 2004.

[22] E. Burger, J. Van Dyke, A. Spitzer, "RFC 4240, Basic Network Media Services with SIP", December 2005.

[23] E. Shi, I. Stoica, D. Andersen, and A. Perrig. OverDoSe: A Generic DDoS Protection Service Using an Overlay Network.Technical Report CMU-CS-06-114, Carnegie Mellon University Computer Science Department, 2006.

[24] L. Subramanian, I. Stoica, H. Balakrishnan, and R. Katz. OverQoS: An Overlay based Architecture for Enhancing Internet QoS. In Proceedings of USENIX, 2004.

[25] David G. Andersen, Hari Balakrishnan, M. Frans Kaashoek, and Robert Morris. Resilient Overlay Networks. In Proceedings of Symposium on Operating Systems Principles, 2001.

[26] J. Liebeherr, M. Nahas, and Weisheng Si. Application-layer multicasting with Delaunay triangulation overlays . Selected Areas in Communications, IEEE Journal on, 20(8):1472–1488, Oct 2002.

[27] S. Gribble, M. Welsh, R. von Behren, E. Brewer, D. Culler, N. Borisov, Steven E. Czerwinski, Ramakrishna Gummadi, Jon R. Hill, Anthony D. Joseph, Randy H. Katz, Z. M. Mao, S. Ross, and Ben Y. Zhao. The Ninja architecture for robust Internet-scale systems and services. Computer Networks, 35(4):473–497, 2001.

[28] J. Liang, X. Gu, and K. Nahrstedt. Self-Configuring Information Management for Large-Scale Service Overlays . In Proceedings of IEEE INFOCOM, 2007.

[29] D. Xu and K. Nahrstedt. Finding service paths in a mediaservice proxy network. In Proceedings of the ACM/SPIE Conference on Multimedia Computing and Networking, 2002.

[30] J. Jin and K. Nahrstedt. Source-Based QoS Service Routing in Distributed Service Networks. In Proceedings of IEEE International Conference on Communications, 2004.

[31] J. Liang and K. Nahrstedt. Service composition for advanced multimedia applications. In Proceedings of SPIE/ACM Multimedia Computing and Networking Conference (MMCN), 2005.

[32] R. Keller. Self-Configuring Services for Extensible Networks – A Routing-Integrated Approach. PhD thesis, Computer Engineering and Networks Laboratory (TIK) ETH Zurich, 2004.

[33] X. Gu, K. Nahrstedt, and B. Yu. SpiderNet: An Integrated Peer-to-Peer Service Composition Framework. In Proceedings of 13th IEEE International Symposium on High Performance Distributed Computing (HPDC), 2004.

[34] B. Raman, S. Agarwal, Y. Chen, M. Caesar, W. Cui, P. Johansson, K. Lai, K. Lavian, S. Machira ju, Z. Mao, G. Porter, T. Roscoe, and Mukund. The SAHARA Model for Service Composition Across Multiple Providers. In Proceedings of International Conference on Pervasive Computing Zurich Switzerland, 2002.

[35] S. Ponnekanti and A. Fox. SWORD: A Developer Toolkit for Web Service Composition. In Proceedings of 11th World Wide Web Conference (WWW2002 Web Engineering Track), 2002.

[36] B. Mathieu, M. Song, A. Galis, L. Cheng, K. Jean, R. Ocampo, M. Brunner, M. Stiemerling, and M. Cassini. Self-Management of Context-Aware Overlay Networks for Ambient Networks. In Proceedings of 10th IFIP/IEEE International Symposium on Integrated Network Management, 2007

[37] X. Gu and K. Nahrstedt. A scalable QoS-aware service aggregation model for peer-to-peer computing grids. In Proceedings of the 11th EEE International Symposium on High Performance Distributed Computing (HPDC), 2002.

[38] X. Gu, K. Nahrstedt, R. Chang, and C. Ward. QoS-Assured Service Composition in Managed Service Overlay Networks. In Proceedings of IEEE 23rd International Conference on Distributed Computing Systems (ICDCS), 2003.

[39] X. Gu and K. Nahrstedt. Distributed multimedia service composition with statistical QoS assurances. Multimedia, IEEE Transactions on, 8(1):141–151, Feb. 2006.

[40] M. Wang, B. Li, and Z. Li. sFlow: Towards Resource-Efficient and Agile Service Federation in Service Overlay Networks. In Proceedings of the 24th International Conference on Distributed Computing Systems (ICDCS'04), 2004.

[41] S. Yamaoka, T. Sun, M. Tamai, K. Yasumoto, N. Shibata, and M. Ito. ResourceAware Service Composition for Video Multicast to Heterogeneous Mobile Users. In Proceedings of 1st International Workshop on Multimedia Service Composition, 2005.

[42] S. Choi, J. Turner, and T. Wolf. Configuring Sessions in Programmable Networks. In Proceedings of IEEE INFOCOM, 2001.

[43] B. Raman and R. Katz. Load Balancing and Stability Issues in Algorithms for Service Composition. In Proceedings of IEEE INFOCOM, 2003.

[44] X. Gu and K. Nahrstedt. Dynamic QoS-Aware Multimedia Service Configuration in Ubiquitous Computing Environments. In Proceedings of IEEE 22nd International Conference on Distributed Computing Systems(ICDCS), 2002.

[45] M. R. Garecy and D.S. Johnson. Computers and Intractability: A Guide to the Theory of NP-Completeness. Freeman, San Francisco, 1979.