ETSI TR 102 805-2 V1.1.1 (2009-11)

Technical Report

User Group;
End-to-end QoS management at the Network Interfaces;
Part 2: Control and management planes solution -
QoS continuity

Reference
DTR/USER-00029-2

Keywords
control, interface, management, QoS, signalling

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00   Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

Individual copies of the present document can be downloaded from:
http://www.etsi.org

The present document may be made available in more than one electronic version or in print. In any case of existing or
perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF).
In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive
within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status.
Information on the current status of this and other ETSI documents is available at
http://portal.etsi.org/tb/status/status.asp

If you find errors in the present document, please send your comment to one of the following services:
http://portal.etsi.org/chaircor/ETSI_support.asp

Copyright Notification

ETSI

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://webapp.etsi.org/IPR/home.asp).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Report (TR) has been produced by ETSI User Group (USER).

The present document is part 2 of a multi-part deliverable. Full details of the entire series can be found in part 1 [i.1].

1 Scope

The present document provides a study of exchange feasibility of the user-related QoS information aiming at E2E QoS continuity.

2 References

References are either specific (identified by date of publication and/or edition number or version number) or non-specific.

- For a specific reference, subsequent revisions do not apply.

- Non-specific reference may be made only to a complete document or a part thereof and only in the following cases:

  - if it is accepted that it will be possible to use all future changes of the referenced document for the purposes of the referring document;

  - for informative references.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication ETSI cannot guarantee their long term validity.

2.1 Normative references

The following referenced documents are indispensable for the application of the present document. For dated references, only the edition cited applies. For non-specific references, the latest edition of the referenced document (including any amendments) applies.

Not applicable.

2.2 Informative references

The following referenced documents are not essential to the use of the present document but they assist the user with regard to a particular subject area**.** For non-specific references, the latest version of the referenced document (including any amendments) applies.

[i.1] ETSI TR 102 805-1 (V1.1.1): "User Group; End-to-end QoS management at the Network Interfaces; Part 1: User's E2E QoS - Analysis of the NGN interfaces (user case)".

[i.2] IETF RFC 3588: "Diameter Base Protocol".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

AmbientGrid: information inference (AmbientGrid) based on the profiles' matching, to structure with grid covering the needed user centric environment

capable QoS: level of QoS that the provider is able to provide

demanded QoS: assertion of the quality level requested by the user

desirable QoS: level of QoS required by user for his service

infosphere: decisional knowledge base managing, in real time, all the personalization and ambient environment information.

offered QoS: assertion of QoS level that the provider proposes to provide

perceived QoS: level of quality experienced by the user

provided QoS: level of quality that the provider has agreed to make available to the user

session mobility: ability to keep the continuity of a service regardless the mobility of the terminal, of the access network, of the core network or of any service components as well as the Service Provider

user-centric session: period of communication between one user and another or other users or servers characterized by a starting time and a termination time, including setting up the relation of the user equipment, access network, core network and services

user mobility: ability for a subscriber to move to different physical locations and be able to use one or more devices connected to one or more access networks to gain access to their services without interruption

userware: innovative user centric middleware (userware) enhancing the seamless feasibility along with the location and activity, personalization and user's ambient contexts

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

3GPP	The 3rd Generation Partnership Project
AF	Application Function
AS	Application Server
AVP	Attribute-Value-Pairs
A-RACF	Access Resource and Admission Control Function
BGS	Border Gateway Services
CCR	Credit-Control Request
CPU	Central Processing Unit
DIAMETER	AAA protocol

NOTE: See RFC 3588 [i.2].

DIFFSERV	Differentiated services (IETF)
E2E	End-to-End
ETSI	European Telecommunications Standards Institute
GPRS	General Package Radio Service
HSS	Home Subscriber Server
IETF	Internet Engineering Task Force
IMS	IP based Multimedia Subsystem
INTSERV	Integrated Services (IETF)
IP-CAN	IP Connectivity Access Network
IPTV	Internet Protocol TeleVision
MBB	Make Before Break
MCF	Media Control Function
MDF	Media Delivery Function
NA(P)T	Network Address (and Port) Translation
NA(P)T-PT	Network Address Translation - Protocol Translation
NGN	Next Generation Network
NSIS	Next Steps In Signalling
NSLP	NSIS Signalling Layer Protocols
PAN	Personal area network
P-CSCF	Proxy CSCF
PCRF	Policy and Charging Rules Function

PDP	Packet Data Protocol
QoS	Quality of Service
QSPEC	QoS Specification
RACF	Resource and Admission Control Function
RACS	Resource and Admission Control Subsystem
RCEF	Resource Control Enforcement Function
S-CSCF	Server - Call Session Control Function
SDF	Service Discovery Function
SDF	Service Data Flow
SDP	Session Description Protocol
SIP	Session Initiation Protocol
SLA	Service Level Agreement
SPDF	Service Policy Decision Function
SSF	Service Selection Function
UE	User Equipment
UMTS	Universal Mobile Telecommunications Systems
UPSF	User Profile Server Function
VPSN	Virtual Personal Service Network
VSC	Virtual Service Community

4 User interaction: existing limitation and user oriented dynamic management

Nowadays, end users desire to access their services, without interruptions and in a continuous way. Moreover, it has to be taken into account dynamic adaptation of services and management of user-centric session according to user preferences and different types of mobility. In this clause, an analysis is carried out on insufficiencies of existing works that are mainly of informational order. In fact, a complete information is needed at the moment of subscription (clause 4.1). The signalling would be more efficient if it has the dynamic information (clause 4.2). And the management would be more flexible if the information IS coherent between the cross layers (clause 4.3). Finally, a user-centric oriented dynamic management (clause 4.4) is defined.

4.1 Information and subscription

User Profile Server Function (UPSF) defined in ETSI and HSS defined in 3GPP store subscribed user's profile information related to one or more service control subsystems and applications. But they do not contain complete profile information such as connectivity subscriptions and PAN (Personal area network) information. In order to offer services adapted to user preferences, the pertinent resource information and user's preference information are needed with the visibility of terminal device, network and service to choose the adequate component (terminal, network, service) to use in a user-centric session. Through interaction with such user system, the pertinent user's data can be reached and maintained.

4.2 Information and signalling

The QoS NSLP proposed by NSIS work group in IETF provides flexibility on patterns of signalling messages that are exchanged. Various QoS models can be used in the network, but these do not affect the specification of the QoS NSLP protocol. The QSPEC carries a collection of objects that can describe QoS specifications in a number of different ways, namely QoS Desired, QoS Available, QoS Reserved and Minimum QoS. A generic template contains object formats for the QoS description, which is designed to ensure interoperability when using the basic set of objects. NSIS has been focused on developing a protocol to manipulate QoS states of network resource along the data path in the network. Nevertheless their work on QoS description does not cover service layer with which user interacts more directly.

To circulate QoS information in a session, SIP has been designed in conformance with the Internet model in the control plane. As far as QoS is concerned, SIP uses SDP to describe the media in the session and negotiate QoS requested. Moreover, SIP can filter information according to the User Profile to implement application servers before establishment of a session. However, is it able to describe the behaviour of service component and communicate the QoS state information between the components in order to re-provisioning services during an active session in a mobility environment according to the user's preferences.

A *dynamic* management of resource information and user's preference information is needed to adapt component (terminal, network, service) during a user-centric session.

For the service continuity in an E2E session, the signalling should support broadening personalisation coverage with QoS through composition of service in service levels.

4.3   Information and management

In the management plane, protocol Diameter is used to carry the information of authorization, policy control and charging etc. The QoS related information can also be mapped to Diameter AVP.

Indeed, if the user wants always to have the dynamic selection of connection and accessibility by any of its identifiers, to have the most appropriate service components in the session with different terminals and different bearers according to different of SLAs, a mechanism is needed to ensure a coherent management of the user information including user's preference between each level (cross-layer). This mechanism could collaborate with E2E signalling.

4.4   User centric approach: dynamic management (convergence of control and management planes)

Generally, the lifecycle of QoS management consists of QoS conception, resource provisioning before the service delivery, transfer of the service data, and management. QoS conception consists of analysis of the system's QoS need and QoS context in order to make available the QoS-related structured information with its access interface. QoS provisioning performed before service delivery covers QoS negotiation which is responsible for issuing an agreement between the components to support the QoS involved in the service, admission control which is used to compare the required resource for the component with the available resource in the system, and resource reservation which arranges for the allocation of resources in response to the user requirements and the agreed QoS levels. When provisioning is achieved, management related functions begin with service delivery in a static way. They analyse the tracked QoS achieved actually and compare it to the initial requirements or agreements.

QoS management aims at satisfying user QoS expectation on the basis of a set of functions. Thus it is important to determine at what time should QoS management be considered and what actions to be performed in the different stages of the service lifetime. To allow user-centric oriented service to be carried out properly, QoS and user's expectation has to be considered for each component not only before but also during the service delivery. That includes the E2E QoS instrumentation in the service design phase and the E2E QoS management in the service operational phase. Furthermore, the management information should be taken into account in the process of re-negotiation.

To achieve the E2E QoS, the provisioning (control plane) and management (management plane) are integrated into the process of transfer the service data (User plane), as shown in Figure 1. During the real-time provisioning, the information management is taken into account during the phase of transfer. This is called dynamic management. Once the service is initiated and during its lifetime, dynamic QoS management is performed in order to ensure that the agreed levels of QoS will be maintained. When QoS degradation is signalled and there is no mean to adjust local resource, a re-negotiation process may be initiated. Dynamic QoS management should be done at every network interface (Home network, Access network, Core network and Service network). Meanwhile, in the phase conception, the instrumentation will be done on all the actors in the session chain, which are user's device, access network, core network, and service components.
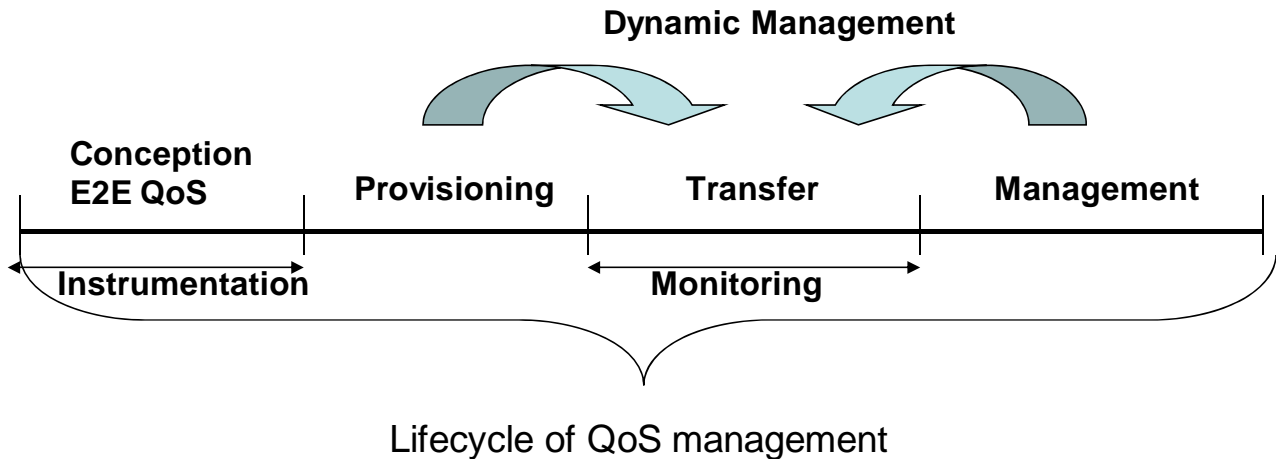
Dynamic Management

Lifecycle of QoS management

Figure 1: Dynamic management

Different solutions could be considered to achieve the E2E QoS dynamic management: a signalling based solution (such as SIP/SDP, NSIS/QSPEC) or a management based solution (such as DIAMETER or Policy Control) can be considered, but for a dynamic management as shown in the Figure 1, the convergence of these two solutions should be studied.

# 5   User-centric oriented QoS signalling

In order to fulfil the needs of QoS E2E negotiation in the user-centric session, a more flexible signalling in higher level to support delivery and negotiation of QoS and user's information between the service components for subscribers across any mobile or fixed network with any user's equipments is appreciated. Based on the VPSN architecture and a dynamic management concept, a "dynamic E2E QoS signalling" in the service level is described which covers the end-to-end session in order to achieve the provisioning for the services demanded and to conform to the SLA. This E2E QoS signalling is able to circulate the description of media as well as the capacity of service component. Moreover, it sends the monitoring QoS state (In/out contract) in the signalling control message during the phase of exploitation. In this clause, firstly the E2E QoS support (VPSN) is presented (clause 5.1), then the E2E QoS provisioning is detailed (clause 5.2). Finally the dynamic QoS signalling during the exploitation with state chart is analyzed (clause 5.3).

5.1   E2E QoS support: VPSN

An E2E user centric session is built by four levels of elements (terminal, access network, core network and service). In the service layer, the service component conception is applied, even the user's terminal running the applet can be considered as a kind of service component in the chain. These service components compose a Virtual Private Service Network (VPSN) to provide a personalized global service. This network is virtual because of the nature of the applicative resources and service components that are sharable as well as the provided abstraction feature. This network is private because it is the logic of linking the service components for the service requested by a particular customer having specific QoS needs. The aim of the VPSN is to satisfy the end-to-end contracted QoS. It means that, the VPSN nodes as well as the VPSN links should be aware of the QoS that they have to provide (local contracts) and the QoS that is currently being provided (current behaviour).

Meantime, the service components which have equivalent function and equivalent QoS compose the virtual communities of services (VSC). The VSC aims to dynamically fulfil QoS commitments by providing an alternative service component to replace the failed component when any QoS change occurs within the VPSN.

From the user's view, service can be accessed, changed and released directly by the end user; all the elements in access/core and service network should be transparent to the end user. The service network should involve all the visibilities of sub-network nodes under its coverage. As a result of such integration of service network and transport network, as the Figure 2 shows, when the service component changes to adapt to the user's demand or the mobility, the sub-networks can establish a corresponding path for this logic link between the service components. Therefore, QoS signalling applied in service layer with the users can be considered as E2E QoS provisioning. The services personalization depends on the user nomadic information like user localisation or the request moment in the QoS signalling.
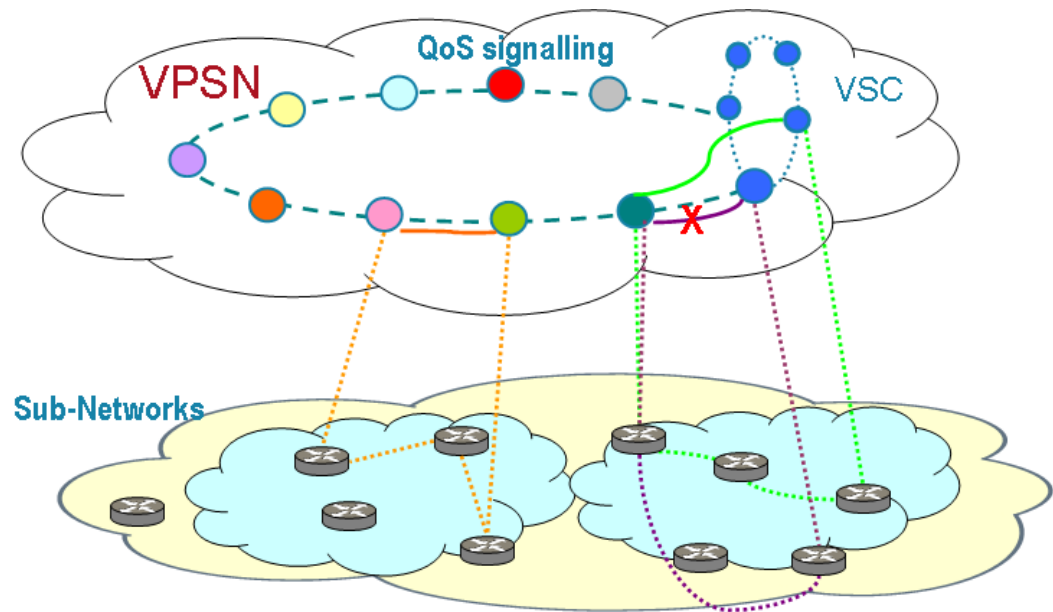
Figure 2: End-to-end QoS

5.2   E2E QoS Provisioning

In each node of VPSN/VSC, is a QoS agent which stores a contracted QoS according to the SLA. The model of expression was refined to distinguish the different values in different views as the Figure 3 shows.
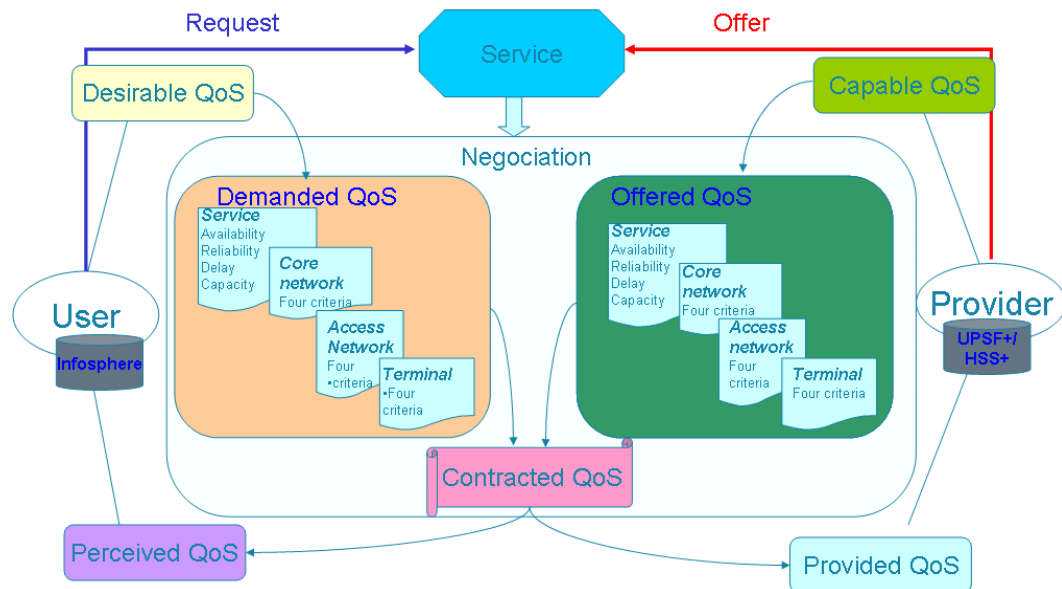
Figure 3: Model of expression for QoS

The Capable QoS is the level of QoS that the provider is able to provide. The Desirable QoS is the level of QoS required by user for his service. The Offered QoS is an assertion of QoS level that the provider proposes to provide. The Demanded QoS is an assertion of the quality level requested by the user. The Provided QoS is the level of quality that the provider has agreed to make available to the user. The Perceived QoS is the level of quality experienced by the user.

The QoS (demanded or offered) is expressed in four levels of visibility according to the four criteria: delay, fidelity, availability and capacity. These criteria can be applied to any QoS classification (Diffserv, Intserv, etc) and can be also easily measurable according to specific parameters. During the service's deployment and provisioning, these four criteria are declined into three categories: conception values, current values and thresholds values.

The conception value is decided at the phase of service conception and capacity planning. It introduces the maximum possibilities of the node's treatments and the link's interactions.

The current QoS is calculated during provisioning and exploitation to reflect the service's behaviour in real time.

The minimum threshold value and maximum threshold value define the range on which the node normally operates. These values are alert thresholds to control the service execution and to announce the problems when the current value exceeds the alerting values.

These values are taken into account during the QoS provisioning. Thus, the service components have the knowledge of the contract to fulfil and the image of its current performance. During the QoS provisioning, if the desired QoS value is below the current value, the service component is activated. On the contrary, if the desired QoS value is more above the current value, the service component will not be activated.

Further, the QoS description of each service component extended into three levels. They are service node QoS, network QoS and equipment QoS, as Figure 4 shows. The service node QoS contains the characteristic of service functions. The network QoS collects the routing table which records the QoS of all the possible paths in the transport layer. If two service's nodes are connected, the real-time QoS condition is calculated and updated in the table. The equipment QoS represents the QoS vision of machine, ex. CPU and memory. A QoS model applicable to these three actors (equipment, network and service node) will enable the aggregation of the End-to-End QoS.
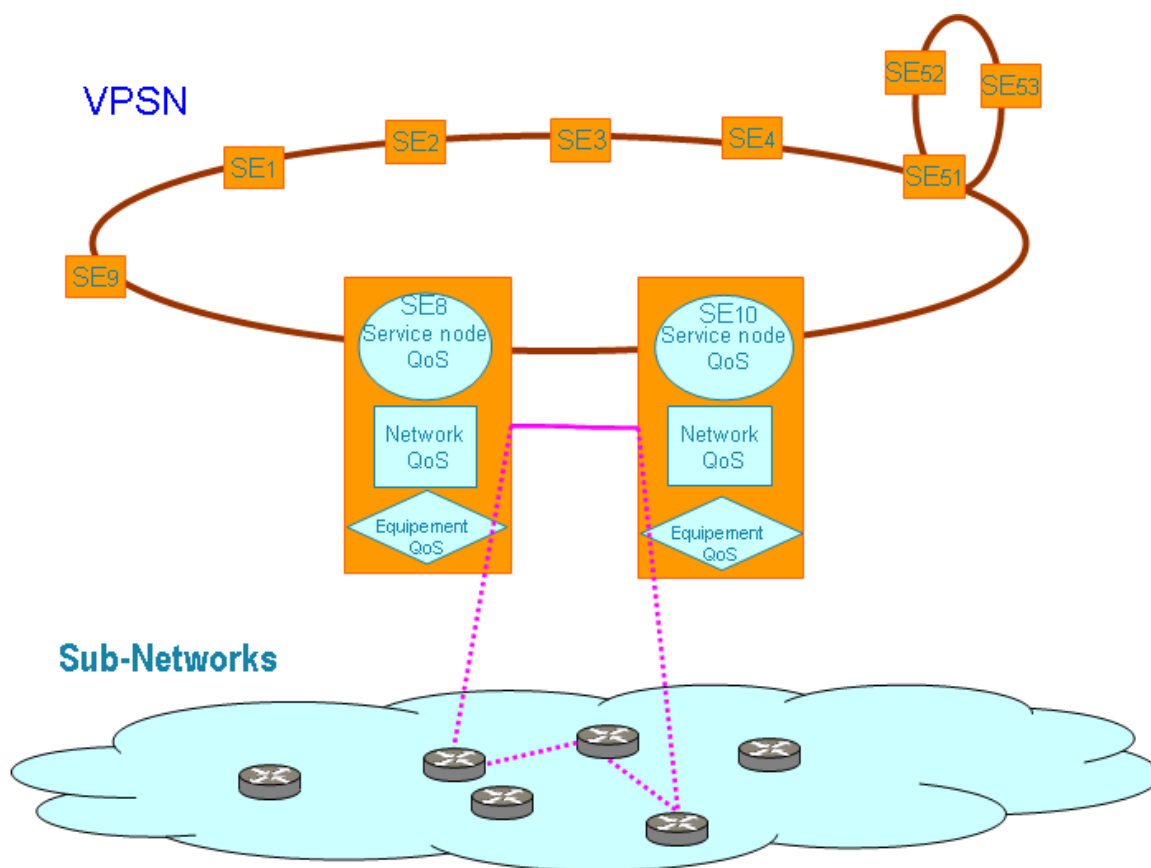

Figure 4: QoS description of service component

5.3 Dynamic QoS signalling during exploitation

The QoS signalling will extend to the process of service delivery, not only before the service delivery. The demanded QoS and Offered QoS (current value) in the node are compared and filtered by the contracted QoS (threshold range values) to get the node's state (In contract/out contract). These states could be transferred in the process of negotiation signalling as to maintain the QoS always conform with the contract dynamically. Since the service management is dynamic, it implies the possibility of corrective actions to various problems during the exploitation of the service. New service components could be added, meanwhile the activated service components could be replaced by the others. In a session opened, when QoS state in one node declines (i.e. the current values exceed the threshold values), firstly the virtual service community (VSC) to which this service component belongs performs a self-management to find another service component with the equivalent QoS to replace. If not, the QoS signalling could interact with user system in the database and re-provisioning the QoS in the virtual personal service network (VPSN) in order to find another link with a QoS compliant to the user's preference. Meanwhile, the sub-network establishes the QoS path simultaneously.

In the user-centric end-to-end QoS signalling, each node could have four states:

IN CONTRACT means the QoS condition is in the scope of contract signed by the user and operator (Current value is in the range of threshold values).

OUT CONTRACT, as the same suggests, means the QoS condition is out of scope of contract signed by the user and operator (current values exceed the range of threshold values). After receiving the first "OUT contract" in the message, the node arms the timer (Timer 1) for waiting the VSC treatment. The state is changed to QoS ADAPTATION.

QoS ADAPTATION is the process of adaptation the contracted QoS. During the period of timer 1, the VSC launches the process of self-management to find another service component to replace. If VSC has not found out one solution till timeout (Time 1), node's state changes to MODIFICATION.

MODIFICATION is the process of negotiation the contracted QoS with the user. The node solicits the database (info sphere & infoware) in order to modify the contracted QoS according the user's preference. The node arms the Timer 2 for this treatment. Until Timer 2 is timeout, the QoS control agent checks the current resource with the updated QoS contract. If the current resource is conforming to the updated QoS contract, the state is changed to "IN CONTRACT". If not, the state is changed to "OUT contract".

Below is the state chart for end-to-end QoS signalling.
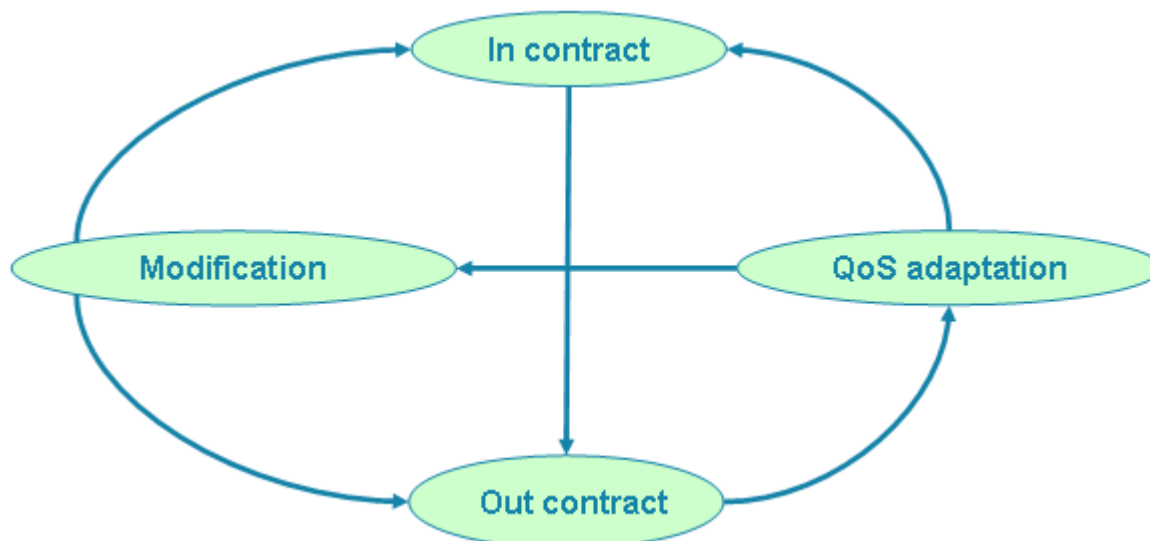
Figure 5: QoS signalling state chart

Besides the four states defined before, the events that cause the change of the state are identified. User Initiated Events (User preference), and Service Initiated Events (self-management of service components in the virtual service community) are external events that trigger the state change of the service node. Meanwhile the QoS condition (In contract/Out contract) and the timer in the entities are needed to notify in the VPSN during the exploitation. The latter are therefore identified as the internal events.

In TR 102 805-1 [i.1], clause 5, the role (initiator, decider and executor) of each important entity and interworking units in the session have been identified. When the mobility occurs, QoS negotiation procedure during the exploitation in the service level uses SIP+ protocol (see below) and the QoS information mapping on other levels (network and user equipment) will be described in clause 6.

The sequence diagrams of two scenarios, user mobility and session mobility, are detailed in the specified use case.

In the scenario below (Figure 6), one session is activated between terminal TE1 and service components SE2, SE3. The end user changes his terminal from TE1 to TE2 (User mobility). For adapting the new terminal's capacity, SE4 is invited by TE2 into the existing session.
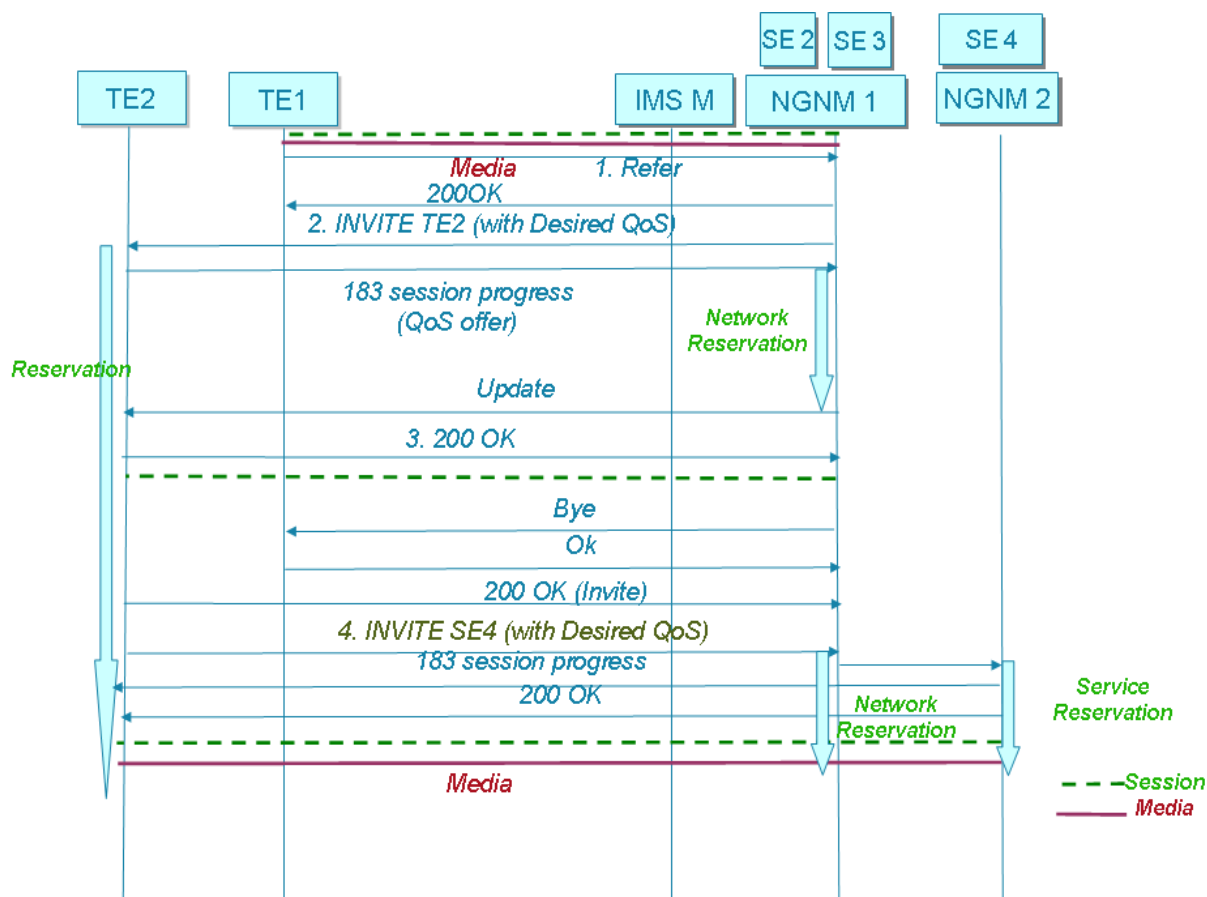
Figure 6: User-initiated update

1)   User launch the trigger to update/modify the session (User changes the terminal during an opened service session), the old terminal sends a REFER message to Server, and Server accept the new terminal.

2)   The Server sends a message INVITE to new terminal (TE2), and provides Service Information in SIP head message and media QoS description in the SDP.

3)   New terminal checks his capability and user profile in the Info sphere, then give a 183 session progress message to Server for informing OK or NOK, and begin to reserve the resource in the networks.

4)   When the new terminal gives a 200 OK response to the NGNM, the TE2 is added into the session.

5)   After the release of the old terminal, the TE2 intend to INVITE the new service element (SE4) for adapting his capacity.

When the resources are reserved, the connection is switched to user terminal.

In the scenario below, the session mobility is illustrated. The SE31 is replaced by SE33 in the session for QoS continuity.

The management system in VSC monitors QoS state of each service element in real-time during the exploitation of service. The management result "IN" or "OUT" Contract could be delivered in the signalling message SIP NOTIFY message from time to time in the VPSN.
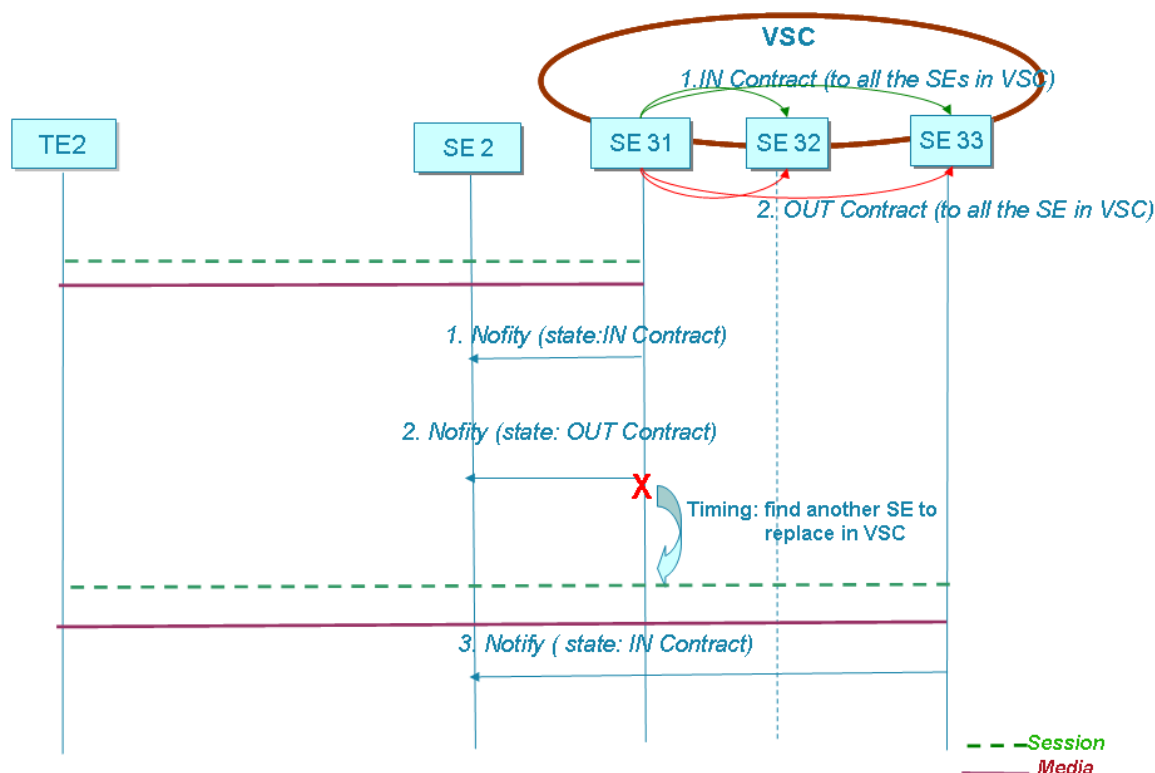
Figure 7: Session mobility related sequence diagram

As Figure 7 shows, the activated SE31 notifies his QoS state to his neighbour elements in same VSC and other service components in the session. When SE31 is "OUT Contract", the VSC find out another SE33 to take the place of SE31 in limit time, The SE33 is activated in the session and notifies his state to SE2. In this way, the service components in one activate session can always keep their QoS state "In contract" according to SLA.

6 User-centric oriented QoS management

After considering the E2E "User-Centric" session QoS signalling in VPSN on horizontal level to meet the needs of the user case representing and concerning the NGN context (Heterogeneity and mobility), in this clause, an E2E session binding carried out on the vertical level is recommended (clause 6.1), followed by related Diameter-based interfaces (clause 6.2) and Diameter AVPs used in the corresponding interfaces to attain E2E user centric session binding (clause 6.3) seems to be the most appropriate.

6.1 E2E session binding

Within this architecture with four levels of visibilities as mentioned in the Figure 8, the mobility in an E2E session can be managed in a temporal dimension.

An E2E session consists of terminal, bearer in the access and core network, and service server. The E2E session binding is broken down into vertical architecturally binding (Terminal binding, Bearer binding, Service binding and IMS session binding), each binding is responsible for integrating component selected in the chain of session. Terminal binding is an association of the terminals and network, for example the binding of the user's security and QoS attributes with certain terminal to the network. The bearer binding is an association between a service data flow and the IP-CAN bearer (for example: the PDP context in the GPRS) transporting service data flow. The service binding is a process of associating service component's workflows in a chain. The IMS session acts as a session middleware between the transport network and the services, allowing having one common access independent signalling platform between bearer and service for providing multiple services components.

Based on the IMS session binding, an end-to-end user centric session binding notion is specified. E2E user-centric session binding associates a service data flow which is defined in a PCC rule by means of the SDF (Service Data Flow) template or in the HSS by means of the service profile template to the bearer deemed to transport the service data flow, to the AS level, and even up to the user equipment (terminal).

Each part in user-Centric session has possibility of mobility. Thanks to the E2E session binding, all the QoS and user's information in the terminal, bearer, IMS session and service components levels can be coherent. Therefore, the user-centric session itself can adapt to any changes by mobility or the ambient environment.
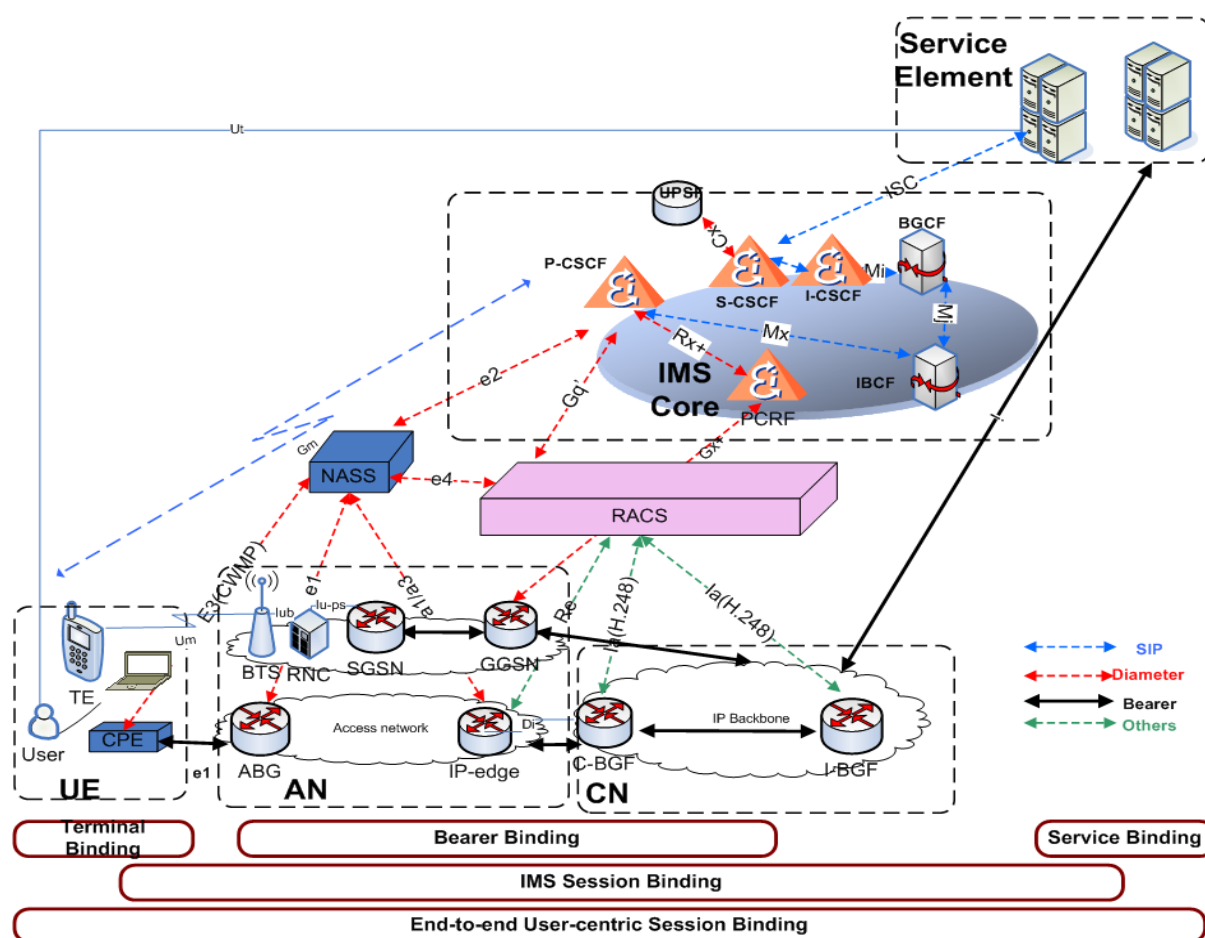
Figure 8: End-to-end user-centric session binding

To meet the functionality required to build and maintain the session "User-Centric", an automate (Figure 9) is specified to take into account all possible events that invoke the binding or re-binding of the "User-Centric" session. Before initiation of a global service for a user-centric session, management of the session is expected to wait for user information and binding information for resource provisioning with the help of Infoware.

The QoS information can be obtained and updated through the Infoware. The user preferences can be taken into account through the "Infosphere".

Once the provisioning finished, one User-Centric session is then obtained. The goal is to maintain a continuous session in real time even if there is any trigger, such as mobility and QoS state change. The change in the geographic location and change of user's activities are external events that trigger the change of the session. The service element's QoS condition change is therefore identified as an internal event. At this moment, management system looks for other QoS suitable service element to replace the failed one in order to maintaining session continuity considering user's preferences. The E2E session binding is desirable to be able to correlate the QoS and user information of each layer in this action. In this fashion, when these trigger event mentioned occurs in any level of session, the adjacent levels should be aware of and self-adapt to any change in mobility and heterogeneity environment so as to realize the dynamic QoS management wished.

Any change in session should be transparent to be seamless, requiring a way of MBB (Make Before Break). MBB means that before release and remove the old component, the session should be already in a relationship with the new component in the chain. The MBB is designed to ensure maximum transparency to the user.
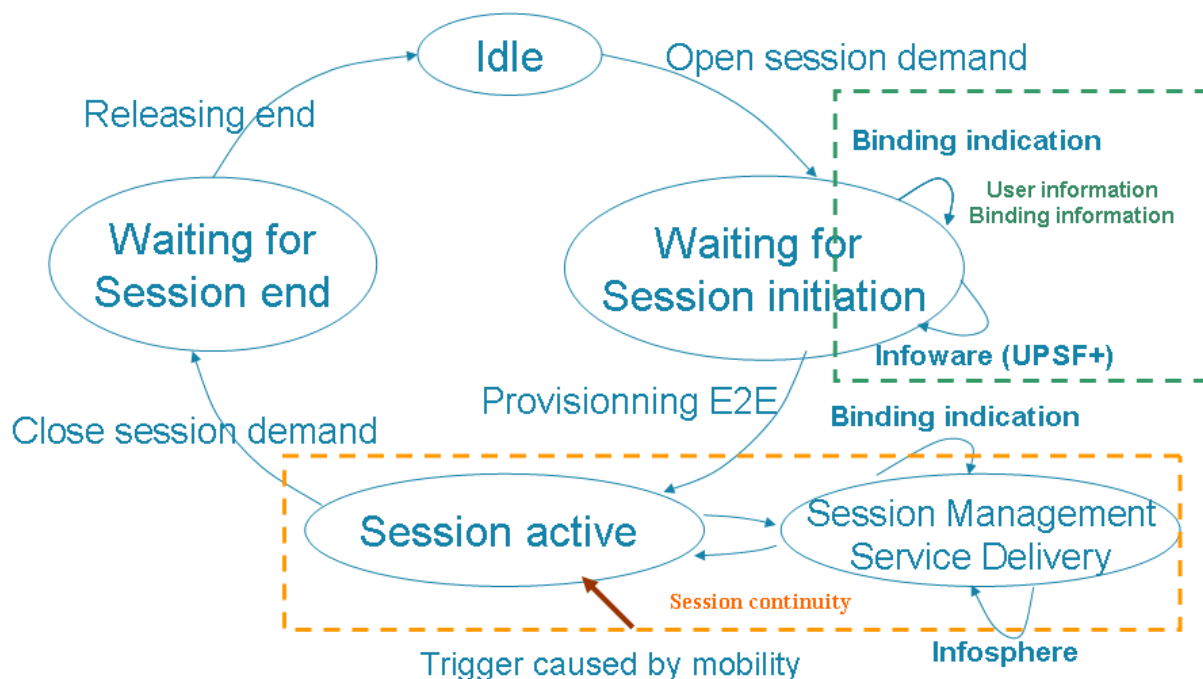
Figure 9: Session binding state chart

An example of E2E session binding in the GPRS network is given in Figure 10. The four entities are arranged, according to the network layer system from top to down, as service session, IMS session, bearer (PDP context in GPRS, UMTS) and User equipment. Each entity has three levels of information: Entity (N) information, Entity (N-1) information and Entity (N+1). In these three levels, the information for binding in the end-to-end session can be found. Entity (N) information contains all the information of the current level. Entity (N+1) information and Entity (N-1) information contain the related Entity ID and the QoS & Charging conditions for correlating.
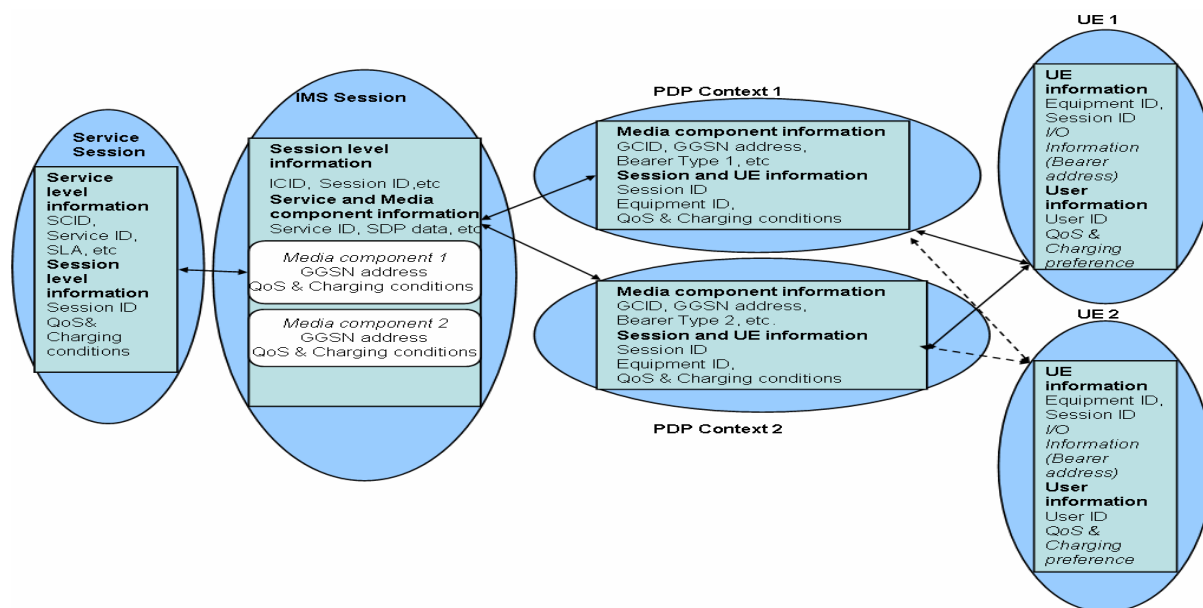
Figure 10: E2E session binding in GPRS network

6.2 Diameter-based interface

Three kinds of information are available in each involved entity:

- QoS information

- Binding indication

- User data information

Reference point Re is located between the RCEF and the RACF. It is used for controlling the L2/L3 traffic policies performed in the transport plane, as requested by the resource management mechanisms, i.e. gating, packet marking, traffic policing and mid-session updates functionalities.

Two pairs of messages are used in this interface are: Credit-Control-Request/Answer is used to indicate bearer or policy rules related events or the termination of the IP CAN bearer, etc. Policy-Install-Request/Answer is used to install the policy rules in the RCEF do the PCC rules provisioning and event triggers for the session, etc.

Reference point Gq' is between the entities P-CSCF (AF) and RACF. It allows the AF to request resources from the RACS. The Gq' enables the transport of NGN Sub-Systems to interact with the RACS for authorization, resource reservation and Border Gateway Services (BGS), i.e. IP filter information to identify the service data flow for policy control, differentiated charging and media/application bandwidth requirements for QoS control.

Four pairs of messages are used in this interface: AA-Request/Answer is used to transfer the session information; Re-Auth-Request/Answer is used to indicate specific re-authorization actions; Session Termination-Request/Answer is used to indicate a session should be terminated. An Abort-Session-Request/Answer is used to indicate that the bearer of the established session is unavailable.

The scenario below shows a service-initiated information update and binding.
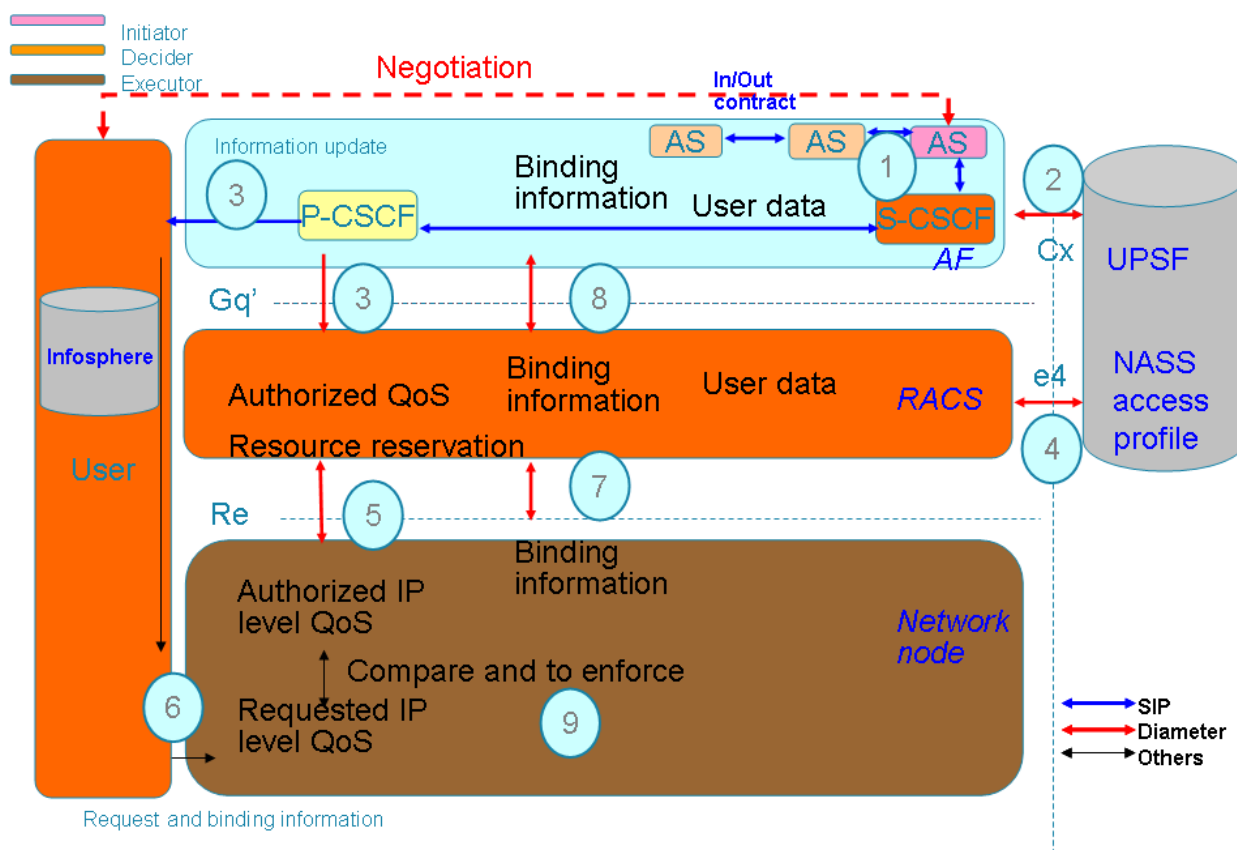
Figure 11: AS-initiated negotiation

1   The AF receives an internal or external trigger to update/modify (OUT contract) the AF session and provided
     Service Information. The AF identifies the Service Information needed (e.g. IP address of the IP flow(s), port
     numbers to be used, information on media types, etc).

2   S-CSCF request the database+ for examining the available resource and user related data in the service layer.

3   The AF provides the Service Information to the RACS by sending a Diameter AAR. Meanwhile, AF provides
     the QoS state change information to user by sending a NOTIFY message.

4   The RACS stores the received Service Information, and requests the user related data containing the allowed
     service, QoS information and user's preference in the database.

5   The RACS identifies the affected established IP-CAN Session(s) using the user data received from database,
     the information received previously from the RCEF and the Service Information received from the AF.

6   UE ensures the mapping of the QoS information in the session to user terminal capability, and sends request
     and binding information according to the user's preference to IP-CAN.

7 to 8 IP-CAN sends the binding information to the Upper layer (PCRF and AF).

9   In IP-CAN, the authorized QoS is compared with the requested QoS resource to enforce.

6.3   AVPs

Diameter based converged interface makes it possible to transfer the QoS and management related information with
homogenous AVPs (Attribute-Value-Pairs) at the same time. Therefore in sub-clause, a look is given into the AVPs
which would be used to attain multimedia service in user centric session with end-to-end session binding.

6.3.1 Multi-service-credit-control AVP

```
DCC (Sub-)Session                    < AVP Header: 456 >
                                     [ Granted-Service-Unit ]
+-----------+                        [ Requested-Service-Unit ]
|           |                       *[ Used-Service-Unit ]
|           |                        [ Tariff-Change-Usage ]
Service-Id a Service-Id b           *[ Service-Identifier ]
      \         /                    [ Rating-Group ]
       \       /                    *[ G-S-U-Pool-Reference ]
        \     /                      [ Validity-Time ]
         \   /                       [ Result-Code ]
          \ /                        [ Final-Unit-Indication ]
         Quota                      *[ AVP ]
           |
           |
       Credit-Pool
```
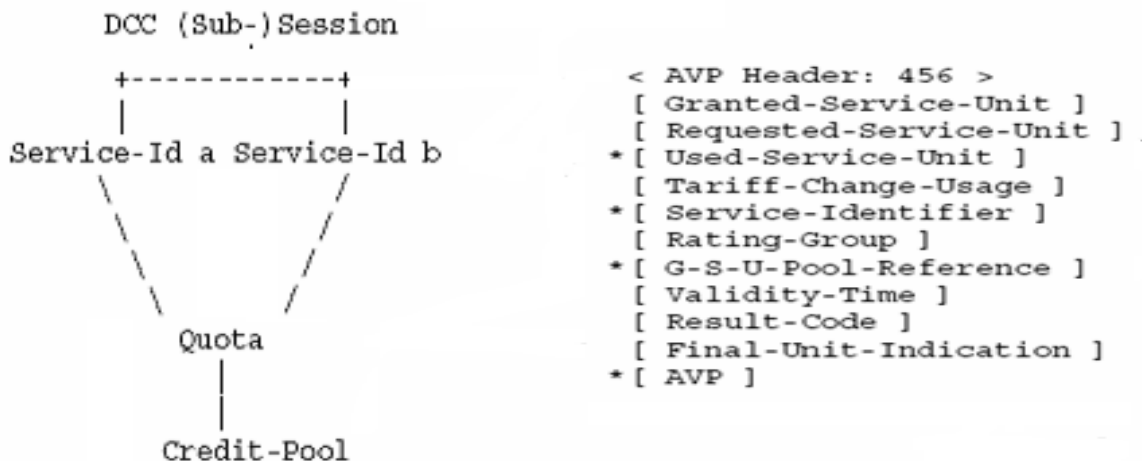
Figure 12: Multi-service-credit-control AVP

Figure 12 shows the different AVPs in the Multi-Service-Credit-Control AVP which would be used for service composition related charging in the CCR message, it indicates the different service IDs and their quotas in the credit-pool. At this moment, Multi-Service-Indicator AVPs should be set as 1 (it means that current SE is capable of handling multiple services independently within a session).

6.3.2 Policy-Rule-Definition AVP

In order to identify the Transport Resources to which the Policy Rule applies, the A-RACF should include at least one of the following Transport Resource Classifiers within the Policy-Rule-Definition AVP in order to associate the policy rule with terminal and network.
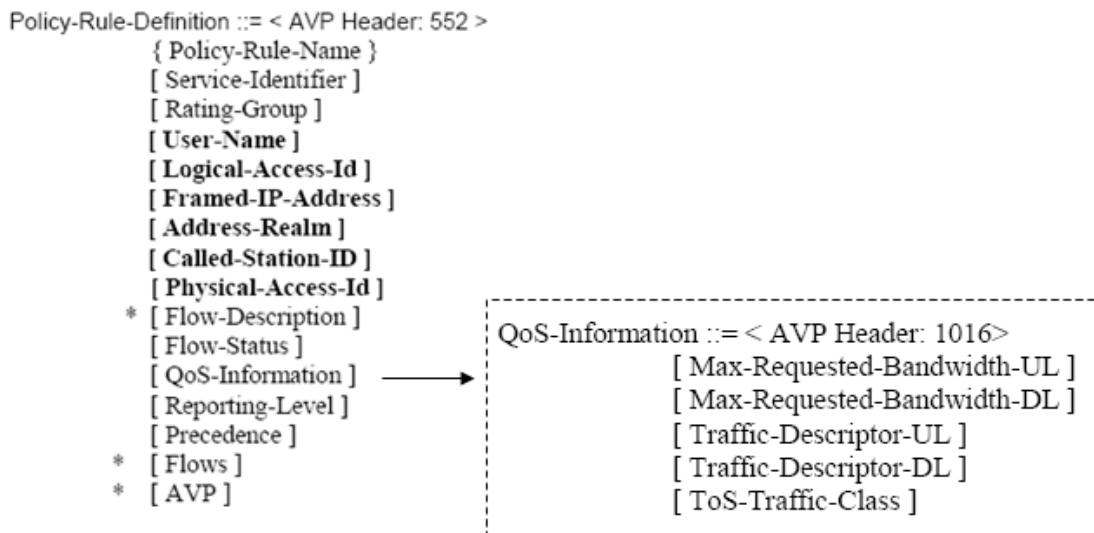
```
Policy-Rule-Definition ::= < AVP Header: 552 >
            { Policy-Rule-Name }
            [ Service-Identifier ]
            [ Rating-Group ]
            [ User-Name ]
            [ Logical-Access-Id ]
            [ Framed-IP-Address ]
            [ Address-Realm ]
            [ Called-Station-ID ]
            [ Physical-Access-Id ]
          * [ Flow-Description ]
            [ Flow-Status ]          QoS-Information ::= < AVP Header: 1016>
            [ QoS-Information ]  ──→            [ Max-Requested-Bandwidth-UL ]
            [ Reporting-Level ]                [ Max-Requested-Bandwidth-DL ]
            [ Precedence ]                     [ Traffic-Descriptor-UL ]
          * [ Flows ]                          [ Traffic-Descriptor-DL ]
          * [ AVP ]                            [ ToS-Traffic-Class ]
```

Figure 13: Policy-Rule-Definition AVP and QoS information AVP

- Logical-Access-Id AVP associates the Policy-Rule-Definition AVP(s) with a particular Logical-Access-Id.

- Physical-Access-Id AVP associates the Policy-Rule-Definition AVP(s) with a particular bearer resource upon which the Policies should be enforced.

- Framed-IP-Address AVP associates the Policy-Rule-Definition AVP(s) with a particular IP Session.

- The Address-Realm AVP may be added if the Framed-IP-Address AVP is also included, in order to associate the Policy-Rule-Definition AVP(s) with a particular Globally Unique IP Address.

- Called-Station-Id AVP associates the Policy-Rule-Definition AVP(s) with a particular Transport Resource on the RCEF.

- The User-Name AVP associates the Policy-Rule-Definition AVP(s) with a particular End User.

- Zero, one or several Flow-Description AVP(s) may be included in the Policy-Rule-Definition AVP, in order to associate a given Policy Rule with IP Flows.

- ToS-Traffic-Class AVP associates the Policy-Rule-Definition AVP(s) with a particular Traffic Class.

6.3.3 Media-Component-Description AVP

```
Media-Component-Description ::= < AVP Header: 517 >
                    { Media-Component-Number };
           *[ Media-Sub-Component ]          Media-Sub-Component ::= < AVP Header: 519 >
           [ AF-Application-Identifier ]              { Flow-Number }      ; Ordinal number of the IP flow
           [ Media-Type ]                          0*2[ Flow-Description ]       ; UL and/or DL
           [ Max-Requested-Bandwidth-UL ]              [ Flow-Status ]
             [ Max-Requested-Bandwidth-DL ]              [ Flow-Usage ]
             [ Flow-Status ]                          [ Max-Requested-Bandwidth-UL ]
             [ RS-Bandwidth ]                            [ Max-Requested-Bandwidth-DL ]
           [ RR-Bandwidth ]
             [ Reservation-Class ]
             [ Reservation-Priority ]
             [ Transport-Class ]
           *[ Codec-Data ]
           *[ Media-Authorization-Context-Id ]
```
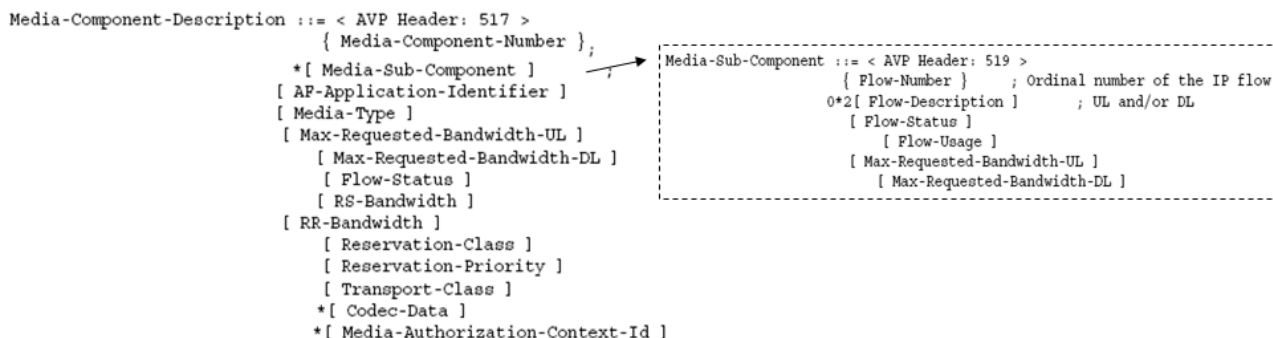
Figure 14: Media-Component-Description AVP

The Media-Component-Description AVP (AVP code 517) contains service information for a single media component within a session. It may contain the Flow-Status AVP, which indicates the particular reservation operation to be performed on the media. The Media-Sub-Component AVP may contain the Flow-Status AVP, which indicates the particular reservation operation to be performed on the flow.

6.3.4 Binding information AVP

```
Binding-information ::= < AVP Header: 450 13019>
                    { Binding-Input-List }        ;
                    [ Binding-Output-List ]         ;
```

Figure 15: Binding information AVP

The Binding-Information AVP (AVP code 450) should be sent between the AF and the SPDF in order to convey binding information required for NA(P)T, hosted NA(P)T and NA(P)T-PT control.

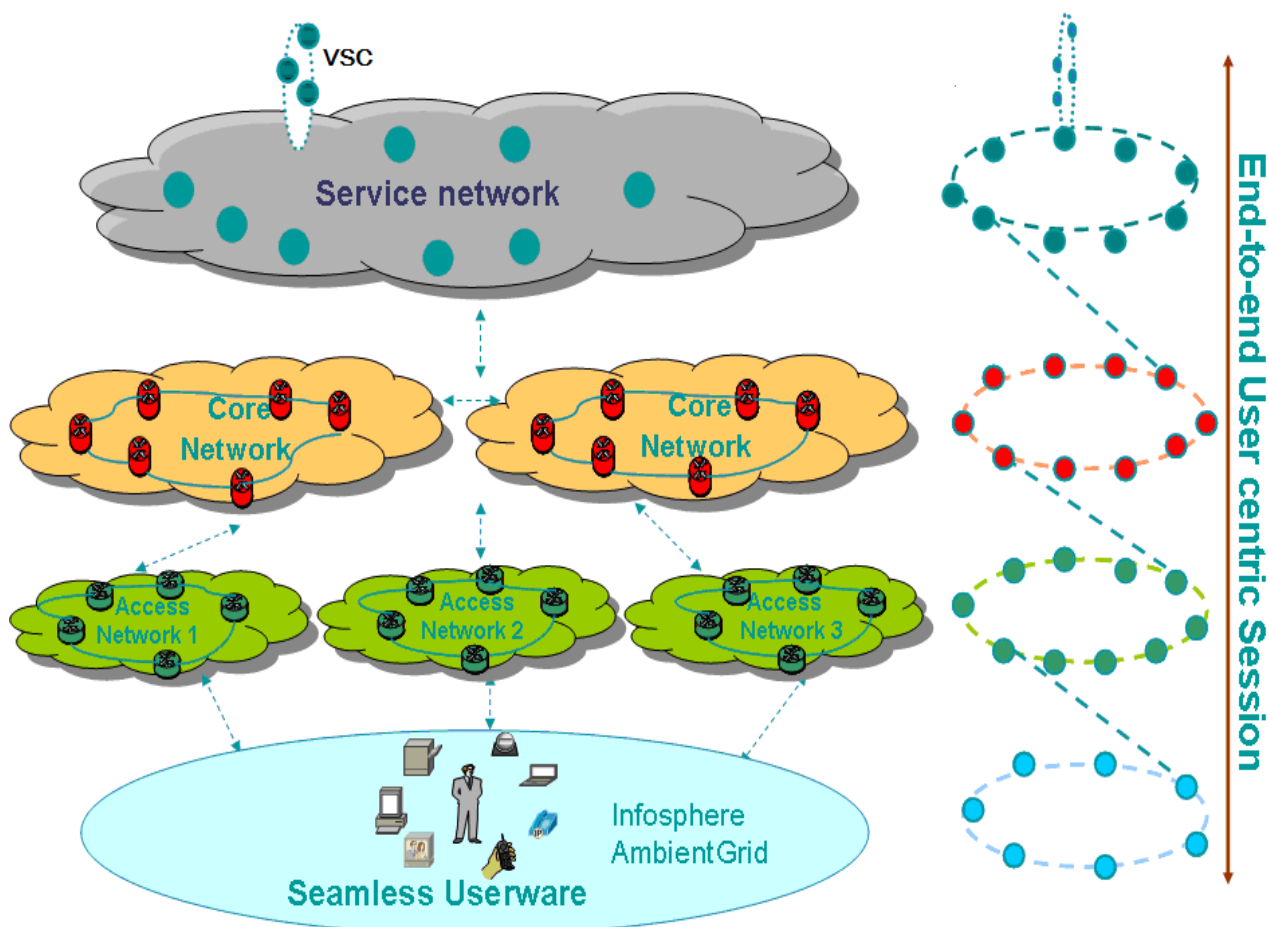# 7 User-centric oriented QoS continuity

Figure 16: User oriented QoS continuity

End users expect to have a continuous comprehensive service throughout the whole session while moving (terminal mobility) or changing terminal (user mobility). During this session, a service is considered as composition of elements in order to adapt any change (session mobility). In such a user-centric approach, in the functional dimension, QoS continuity should cover three basic functions:

- End-to-end provisioning in horizontal architecture

- End-to-end session binding in vertical cross layers

- SLA conformity

In the protocol dimension, the QoS signalling protocol should interact with the information collected in four layers(User and terminal layer, access network layer, core network layer and service layer) and monitoring QoS state in each node at the proper time. It works in the horizontal service layer to provisioning resource for the services respecting demanded QoS, cooperating with the dynamic E2E session binding mechanism.

In the information dimension, the intelligent seamless userware (Infosphere and Ambient Grid) locates in side of user to help him always keep the personalization and ambient resources information with him. Moreover, in operator's side, an enhanced UPSF offers actually an informational inference on four levels of visibility in order to have more dynamicity than the simple UPSF. Thanks to the complete information (user preferences, user equipments, network and service subscription) on the two sides (User and operator), the user-centric session can be chained from the user terminal, the bearer, the IMS core network until service session.

8 User E2E QoS Measurement

One E2E QoS measurement is to trace and monitoring the performance of each step hop by hop (transaction by transaction) from user to service. The service delivery includes the following steps: user network attachment, user service demand, resource reservation, session establishment and media transfer. The different interaction flow among related entities for each step identified in the Figure 17 for example, should be measured in term of transmission delay, fidelity (error rate), and etc, according to the defined QoS metric [i.1]. In addition, an example case of IPTV service for E2E QoS measurement is presented in Annex A.
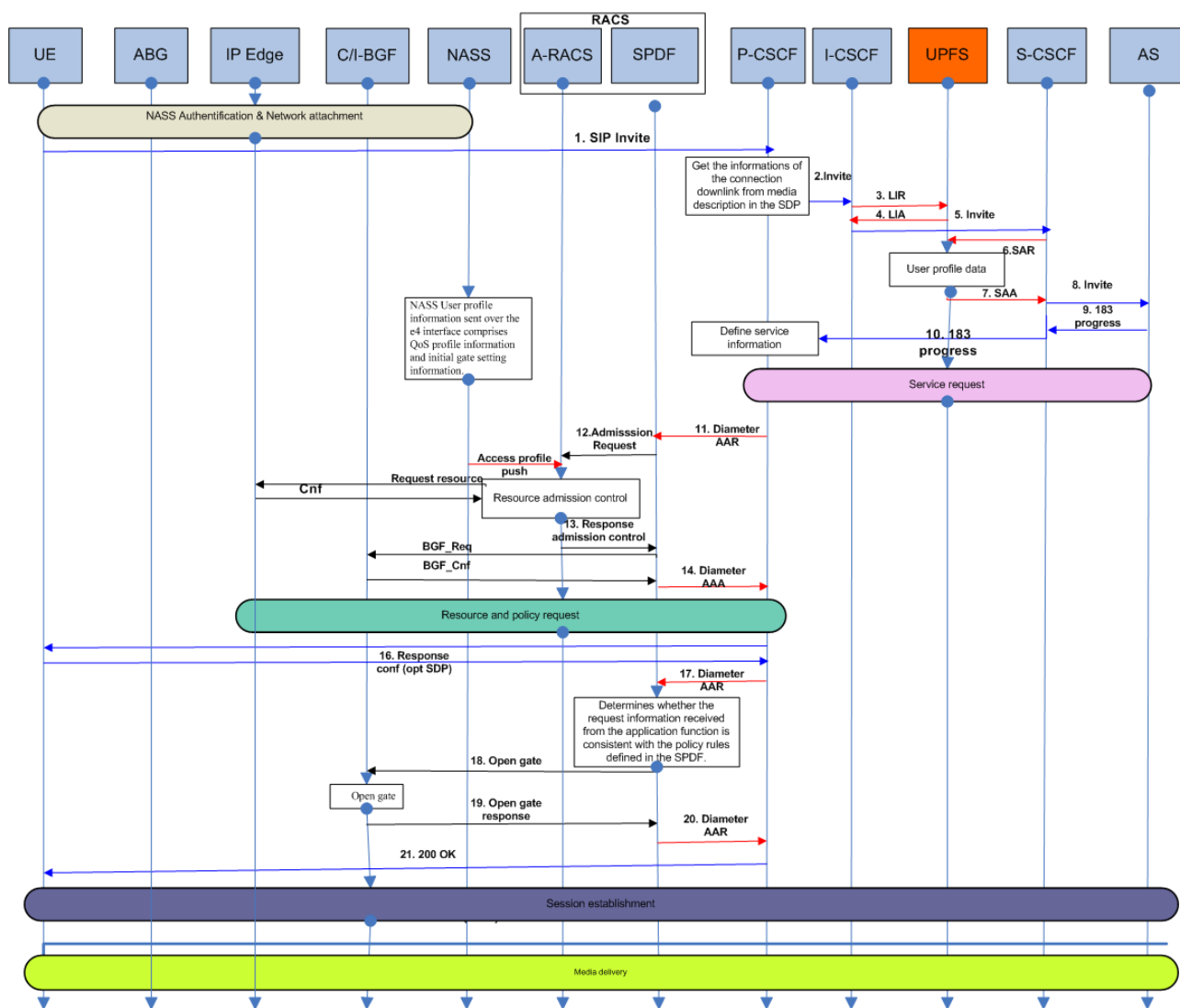

Figure 17: Identification of QoS measurement steps

During the media transfer step, QoS measurement is basically close to network traffic measurement. In fact, there are several media flows in one session opened. The performance characteristics of these different media data flow should be measured hop by hop and analysed in the corresponding measurement point in the network segments separately. The results are stored in the shared knowledge database (Figure 18). However, for end-user, end-to-end QoS (Media delivery) issue is significant. The QoS metric are applicative debit, response time, message error rate, message lost rate, service accessibility rate, etc. QoS metric are measured from user's demand to media received. In these measurements, the QoS metric of platform and terminal functions (two extremities of user session) assuring the media delivery is considered. In the network level, the measurement points locate in the network borders routers, for instance, the entry points, exit points and interworking points etc. These segments are listed in Figure 18.
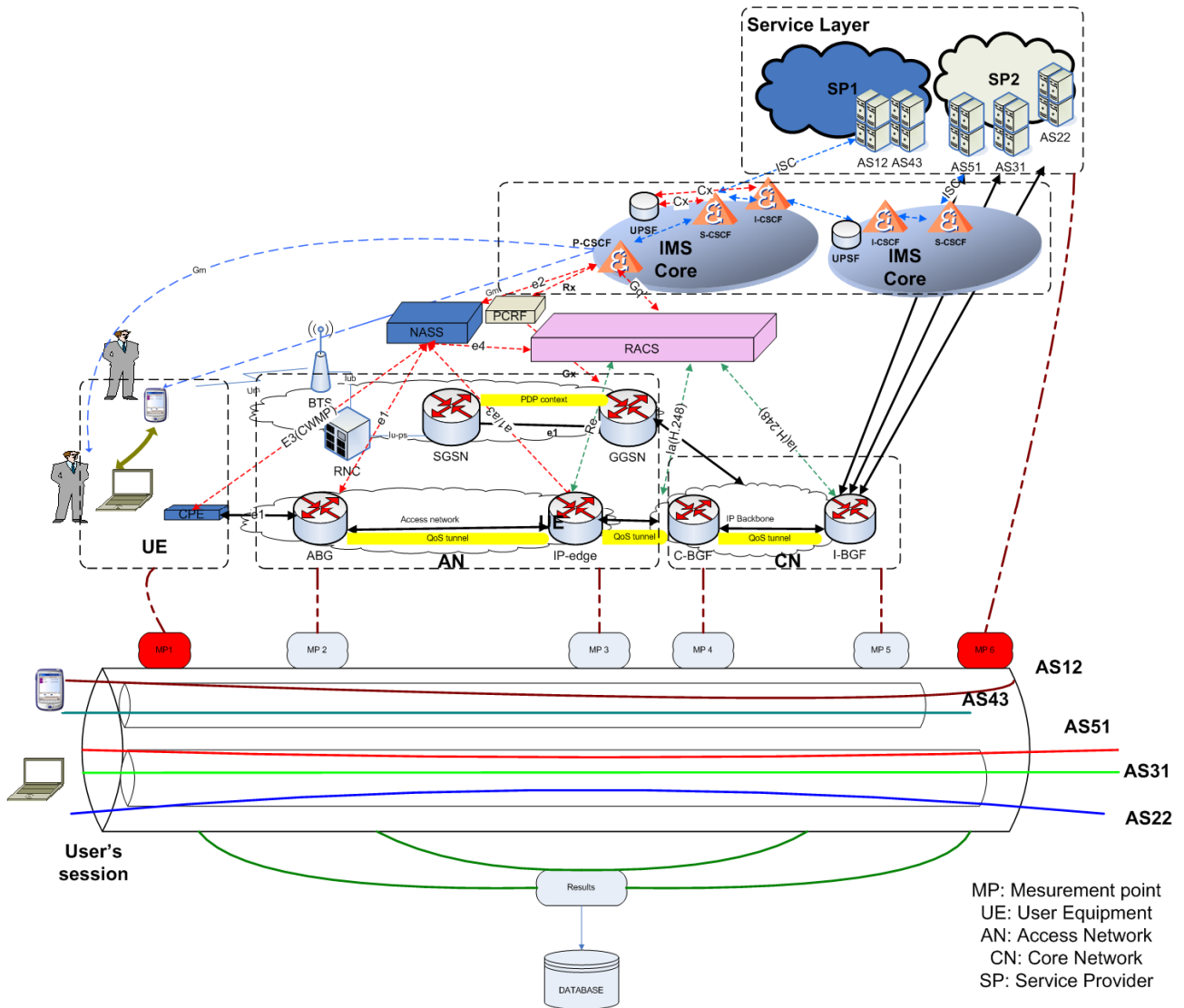
Figure 18: Media delivery QoS measurement

- Service Media delivery: MP1→ MP6/MP6→MP1.

- Service access network: MP1→MP2.

- Service access sub-network: MP2→MP3.

- Interworking service (AN and CN): MP3→ MP4.

- Service core network: MP4→MP5.

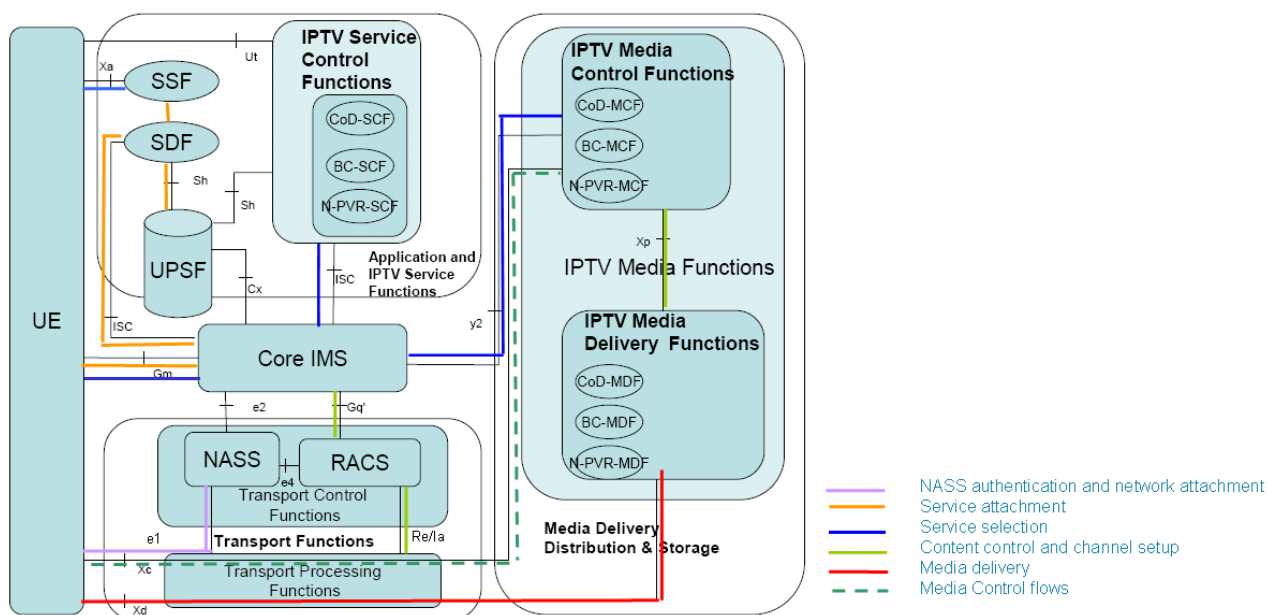- Service access server: MP5→MP6.

Annex A:
IPTV

Figure 19: IPTV Service

IPTV architecture shows clearly the levels of visibility defined in the user case [i.1], which are user equipment (UE), network (bearer and transport control functions), IPTV media functions (MCF and MDF), core IMS and service platform (Application and IPTV service functions),).

IPTV service functions as shown in Figure 19 provide basic integration of Service Control functions, Service Selection Function (SSF) and Service Discovery Function (SDF) and UPSF.

In the case of IPTV service delivery, each corresponding steps for service delivery mentioned in Clause 8 are identified (The corresponding flows are identified by different colour lines.):

1)   NASS authentication and network attachment: interface e1

2)   Service request: In the step of service request mentioned in chapter 8, the end user request to just attach a related Application Server. In IPTV service, firstly the end user attaches to the service platform instead of a specified AS via SIP related interfaces Gm (between User equipment and Core IMS) and ISC interface (between Core IMS and SDF). This phase provides the personalisation service discovery (Services attachment). And then, UE retrieves data related to specific service and makes appropriate service selection via Xa interface (Service selection). The session initiation request is routed by the Core IMS (SIP based interface Gm) entities up to the SCF (SIP based interface ISC) in charge of the requested service.

3)   Resource reservation: From the control entities to transport entities via interfaces Re/Ia and Xp.

4)   Media delivery: Logical interface Xd between UE and MDF.

For the QoS measurement, the corresponding measurement point and different transaction flows in the IPTV service are identified (Figure 20).
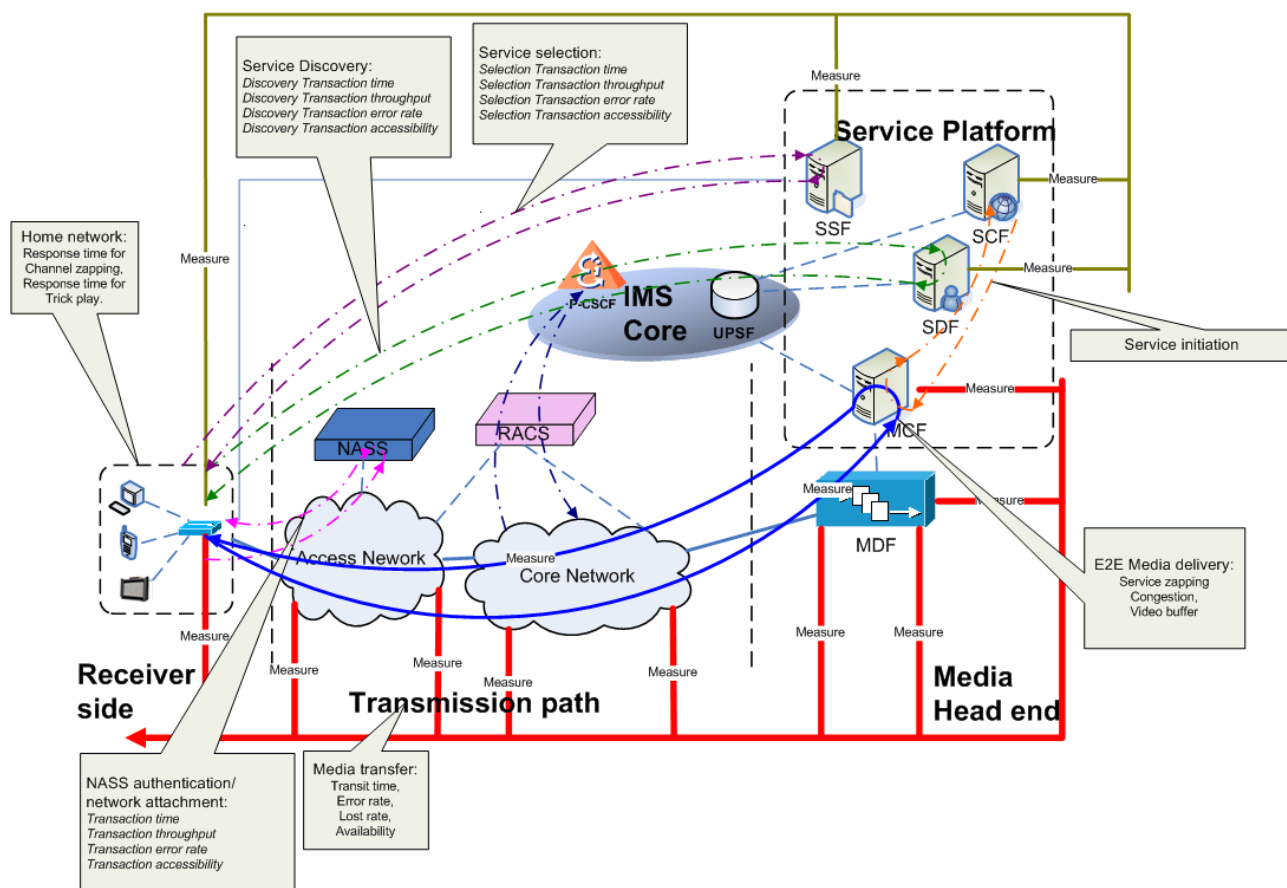
Figure 20: QoS measurement in IPTV

The E2E IPTV Service delivery measurement contains:

- QoS performance (for example: transaction time, throughput, error rate, accessibility) for transaction flows before the media delivery: NASS authentication and network attachment (pink dotted line), service discovery (green dotted line), service selection (violet dotted line), service initiation (orange dotted line), and resource reservation (blue dotted line).

- During media delivery, the aspects below based on QoS model [i.1] should be considered:

    - Service zapping and trick play transaction in E2E media delivery (Blue line). The control performance between MCF and MDF should also be considered in this transaction.

    - QoS in transmission path: Service access network (home networking gateway/User equipment to access network), Service access sub-network, Interworking service (AN and CN), Service core network and Service access server (Network to MDF).

Annex B:
Bibliography

- ETSI TR 102 805-3 V1.1.1: "User Group; End-to-end QoS management at the Network Interfaces; Part 3: QoS informational structure".

- ETSI ES 283 026: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN);Resource and Admission Control; Protocol for QoS reservation information exchange between the Service Policy Decision Function (SPDF) and the Access-Resource and Admission Control Function (A-RACF) in the Resource and Protocol specification".

- ETSI ES 283 034: "Telecommunications and Internet converged Services and Protocols for Advanced Networks (TISPAN); Network Attachment Sub-System (NASS); e4 interface based on the DIAMETER protocol; Release 2".

- ETSI TS 183 017: "Telecommunications and Internet Converged Services and Protocols for Advanced Networking (TISPAN); Resource and Admission Control: DIAMETER protocol for session based policy set-up information exchange between the Application Function (AF) and the Service Policy Decision Function (SPDF); Protocol specification".

- IETF RFC 4006 (August 2005) H.Hakala, L.Mattila, J-P.Koskimem, et al.: "Diameter Credit-Control Application".

History

Document history		
V1.1.1	November 2009	Publication