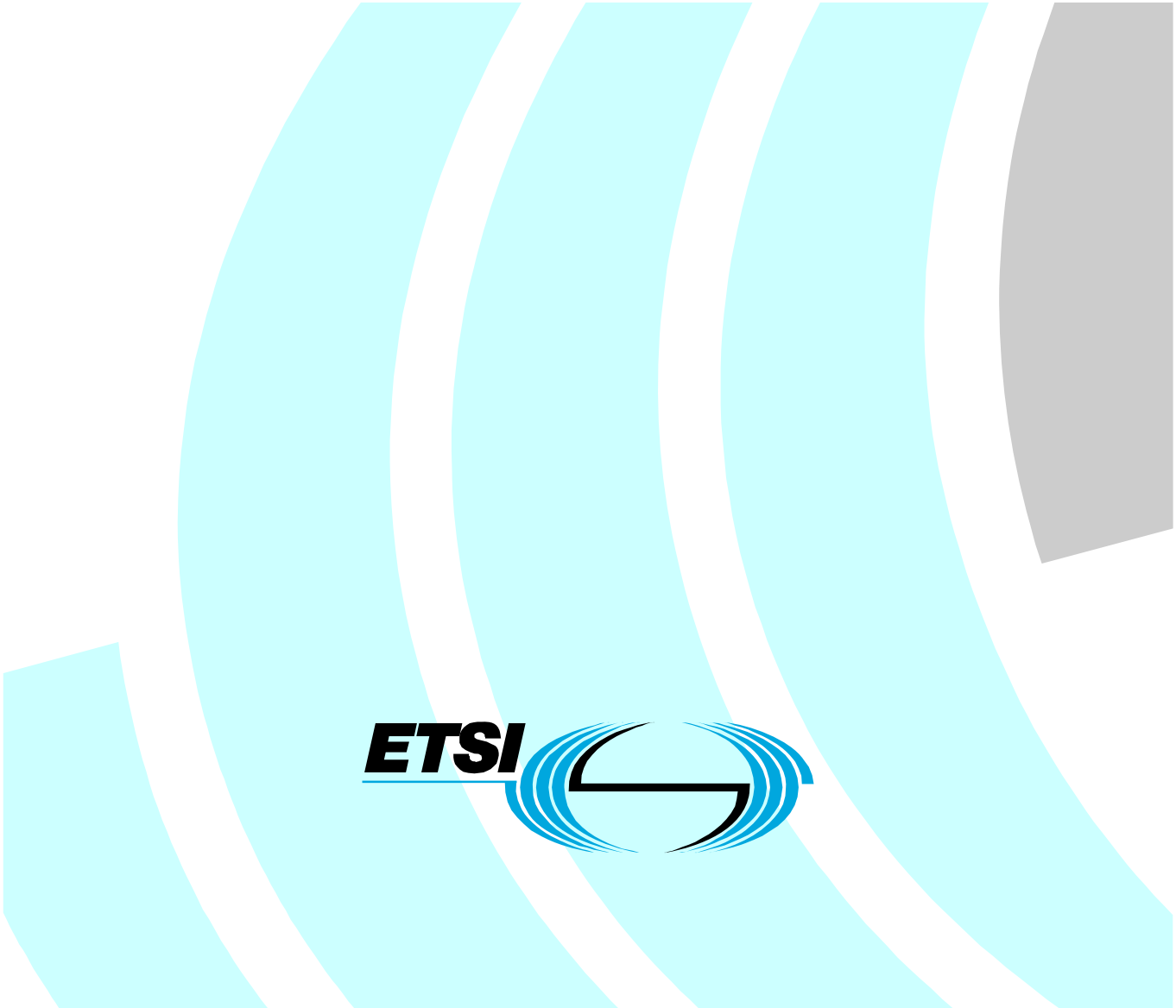
ETSI TS 181 018 V2.0.0 (2007-08)

Technical Specification

Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); Requirements for QoS in a NGN



Reference
DTS/TISPAN-01046-NGN-R2

Keywords
analysis, QoS

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

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 Specification (TS) has been produced by ETSI Technical Committee Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN).

1 Scope

The present document identifies the capabilities that a TISPAN NGN shall provide in order to guarantee an adequate Quality of Service (QoS) to the media flows and it defines the specific requirements for an NGN environment related to the identified capabilities.

The present document defines requirements for a TISPAN Release 2 NGN.

The present document does not define classes of services or values for the parameters used to define the classes of services: this work is done by other standardization bodies.

The present document defines detailed requirements based on clause 4.11 in [1].

NOTE: The present document uses the term "NGN" only in the context of TISPAN.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- 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.
- For a non-specific reference, the latest version applies.

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.

[1] ETSI TS 181 005: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); Services and Capabilities Requirements".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in TR 180 000 and the following apply:

network probe: network element able to intercept the media flow and provide some information to the control layer

QoS reporting: this mechanism identifies the ability for some network elements to collect the values of some QoS metrics of a single service instance

NOTE: Example of QoS metrics could be delay, packet loss, etc.

resource monitoring: this mechanism identifies the ability to monitor the topologies and resources of the transport segments controlled by RACS

NOTE: Resource monitoring includes detecting the actual usage of these resources.

3.2 Abbreviations

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

ADSL	Asymmetrical Digital Subscriber Line
AF	Application Function
ATM	Asynchronous Transfer Mode
BGF	Border Gateway Function
BoD	Bandwidth on Demand
BRAS	Broadband Remote Access Server
CAC	Call Admission Control
CAC	Connection Admission Control
CBR	Constant Bit Rate
CLP	Cell Loss Priority
CoS	Class of Service
CPE	Customer Premises Equipment
NGN	Next Generation Network
OSS	Operating Support Systems
PCR	Peak Cell Rate
QoS	Quality of Service
RACS	Resource Admission Control Subsystem
TRIM	Topology and Resource Information Model
UE	User Equipment
VBR	Variable Bit Rate
VC	Virtual Channel
VP	Virtual Path

4 General requirements

The following general requirements shall be fulfilled in a TISPAN NGN:

- The NGN shall support QoS reporting mechanisms in order to indicate to the control/service/management elements (e.g. RACS, accounting functions) the QoS really achieved by the bearer flows.
- The NGN shall support resource monitoring mechanisms in order to determine the available network resources (e.g. link bandwidth, port utilization, queue depth). The assumption is made that creation of the inventory of the network resources shall be done via suitable interfaces.
- The NGN shall support dynamic policy provisioning mechanisms in order to allow the change on demand of the policies applied to a single user access (e.g. change the maximum bandwidth of an ADSL access).
- The NGN shall support end-to-end QoS control to allow guaranteed QoS levels to be provided between connectivity end points. Several QoS classes should be supported.
- In a single provider domain environment, according to the user service profile, the NGN shall be able to downgrade the QoS after a QoS degradation is detected.

5 Specific requirements

5.1 QoS reporting

5.1.1 Information sources

It is required that at least one of the following elements is able to provide QoS reporting information per service basis:

- CPE/Customer Network Gateway.

- Border Gateway Function (BGF).
- Network probes.

In addition it should be possible to have also other information sources such as Access Nodes, IP edge nodes, etc.

5.1.2 Services

Gathering of QoS reporting information shall be provided for both session and non session based services (e.g. content based services).

5.1.3 QoS reporting information

QoS reporting information:

- For session based services reporting at least the following information shall be provided: number of packets received, packet loss, average delay, and jitter. In addition, it should be possible to also report other quality metrics.
- For non session based services reporting at least the following information should be supported: number of packets received, packet loss, number of retransmitted packets. In addition, it should be possible to also report other quality metrics.

It may also be possible for the information sources to process the information gathered. For example, processing can be used to determine the urgency in reporting the information and/or to reduce the forwarding capacity needed to transfer the information or for generating aggregate information starting from raw ones.

5.1.4 Granularity

The granularity used by the network elements to generate the QoS reporting information shall be an operator choice on a per service basis and per information source basis. The QoS reporting may be triggered by pre-defined events, e.g. thresholds violation, during a session or a service instance. Exhaustive QoS reporting information at the end of service shall be supported (e.g. communication session, video streaming). The level of granularity shall be such that it is sufficient for purpose without causing excessive load to the requester and responder of QoS reporting, as well as the network traffic.

The QoS reporting should contain a timestamp indicating the time at which the observation was made.

5.1.5 Use of QoS Reporting information

Use of the QoS reporting information:

- The control/service/management elements that use the QoS reporting information shall be able to process the information gathered.
- The QoS control layer shall be able to change the admission control and/or enforcement policies used for QoS purposes based on QoS reporting information. The policies to be changed are an operator choice and could be one or more of the following:
    - the enforcement policies for the media flow related to the QoS reporting information;
    - the enforcement policies for other media flows of the same user;
    - the enforcement policies for the media flows of other users;
    - the admission control and/or enforcement policies for a particular network segment;
    - the admission control and/or enforcement policies for new media flows requests.
- The accounting functionalities shall be able to use the QoS reporting information to enrich the CDR provided to the billing systems.

- The management functionalities shall be able to use the QoS reporting information for off line analysis (e.g. for planning purposes).

5.2 Resource monitoring

5.2.1 General

The ability to offer end-to-end QoS control in an NGN is of fundamental importance as new services and applications with ever increasing QoS requirements become available. Management of application QoS requirements where a number of services are being provided over individual links requires knowledge of the available bandwidth on the link.

In order to achieve these requirements the RACS needs to have an accurate and current knowledge of the available network resources in the transport layer, e.g. in case of a network device failure or congestion occurring this may impact on available QoS due to increased load or delays caused by re-routing. The RACS needs also to have knowledge of the status of each network component (nodes and links) under its control.

This information may be held in accordance with a Topology and Resource Information Model (TRIM) and made available to RACS to provide the required information for QoS control. The sources and types of information are described in the following clauses.

In general, RACS shall have knowledge of the network topology, routing information and total resource capacity including capacity in each forwarding class and shall be made aware of dynamic changes in the network (e.g. device or link failure).

5.2.2 Information sources

Topology, resource and routing information may be acquired from a number of different sources, including external systems such as OSS and directly from the network elements controlled by the RACS system. The information may be stored in accordance with the information model and made available to the RACS system.

Resource Monitoring employs a number of different sources: primarily the Access Node and IP Edge, but where true end-to-end QoS provision requires knowledge of the number or type of CPE connections, the CPE/Customer Network Gateway must support the resource monitoring mechanism in order to monitor the whole network from CPE to BGF. The information sources are dependent on the access technologies in use.

Information Sources: sources of the information shall be OSS Functions and whatever resource(s) the RACS controls.

RACS may need information from one or more elements about their environment. Although information must be maintained about the whole network, aggregation may allow RACS to receive the required information from a subset of these elements.

In addition to the above list it may be possible to identify other information sources (e.g. CPE) depending on the type of access network in use. Consideration should be given to a range of access architectures when identifying the possible information sources.

5.2.3 Resource Monitoring information

The Resource monitoring information available may vary depending on the type of resource in use. This in turn depends upon the type of access architecture involved. Some consideration also needs to be given to the types of application in use and their QoS requirements as these have an impact on the resource information required. The information are dynamically updated and stored in such a way as to be accessible to the RACS.

Types of information: can be separated into three different categories:

- Devices and interfaces of the physical network topology.
- Routing information that describes the connectivity through the topology.
- Resource availability information.

Within these categories there may be static and dynamic information. For instance the topology information may comprise static information acquired during the bootstrap phase, plus dynamic updates due to devices or interfaces taken out of service due to failure or for maintenance purposes etc. Routing information is potentially dynamic as it may change due to re-routing caused by network topology changes. Resource availability is initially provided as static information, but is updated dynamically with usage.

Resource monitoring information: the resources monitored shall be:

- Network devices (static).
- Links (static).
- Bandwidth available on the link (static or dynamic).
- Port utilization (dynamic).
- Queue depth (static).
- Queue utilization (dynamic).

In addition it should also be possible to monitor other information such as latency, L2 capacity, port capacity, VP capacity and VC capacity, CoS bandwidth availability where required - see clause 5.2.5.

It may also be possible for the information sources to process the information gathered e.g. to reduce the overhead in the RACS.

In addition RACS may be able to derive some information by processing the available data (e.g. to calculate the current available link bandwidth from the total link bandwidth figure by maintaining knowledge of the current utilization of resources).

5.2.4 Granularity

The granularity for providing the Resource monitoring information shall be a trade-off between the amount of information and the traffic within the network. Scalability issues need to be considered to avoid a high management overhead.

Granularity:

- The resource information shall be made available in a timely manner, e.g. scheduled or periodic.
- The granularity used by the network elements to generate the resource monitoring information shall be an operator choice on a per resource basis and per information source basis.
- The level of granularity shall be such that it is sufficient for purpose without causing excessive load to the resource or monitoring function, as well as the network traffic.
- The resource information should contain a timestamp indicating the time at which the measurement was made

5.2.5 Use of resource monitoring information

The Resource monitoring information may be used for different purposes: primarily to provide the knowledge to the control elements to enable them to make informed decisions on what policies to apply to the network elements, but it could also be used for other purposes such as to monitor performance for fault prediction or other purposes, or for off-line management or planning purposes.

Use of the Resource monitoring information:

- The Resource monitoring information shall be in a suitable format to enable the control/service/management elements that use it to be able to access and process it.
- The RACS shall be able to change the admission control and/or enforcement policies used for QoS purposes based on the Resource monitoring information. This should apply to the admission control and/or enforcement policies for new media flows requests.

- The Resource monitoring information shall be made available to other sub-systems for the purposes of performance/fault management, planning purposes, etc.

5.3 Dynamic policy provisioning

As a key part of an NGN's ability to offer QoS, RACS shall be able to activate and enforce existing policies governing the bandwidth available to a particular User. In addition to this, RACS shall also be able to dynamically define new policies based on information from a number of sources, including network devices and management systems, and implement these policies.

This places a number of requirements on the TISPAN architecture:

1) RACS shall support provisioning and configuration of policies to be used to guarantee the requested QoS level. This includes to dynamically create/update/remove/query/activate/de-activate policies.

2) RACS should be able to control both Upstream and Downstream bandwidth resources.

3) RACS shall be able to change the Upstream and Downstream bandwidth available to a particular subscriber. Such changes may apply to all traffic classes offered to the subscriber or a subset of those classes.

5.4 End-to-end QoS requirements

When a user requests a service it is likely that a number of different network domains are involved in providing the service to the user. If the service is offered with a given level of QoS there must be a mechanism to pass the QoS requirements over the end-to-end path to ensure that this level of QoS is maintained. The process of end-to-end QoS control can be divided into two phases:

1) The QoS requirements are passed to the relevant network domains.

2) The network domains act on the requirements to enforce the requested level of QoS on the specific end-to-end connectivity path.

Optionally a third phase could be to monitor the connectivity path to ensure the requested QoS is maintained.

Requirements

- The NGN shall provide a mechanism to pass per session or aggregate QoS requirements to each relevant network domain. QoS requirements may be different for the signalling and media flow paths.

- The NGN shall support a mechanism to apply the end-to-end QoS requirements to the transport layer in each domain between connectivity end points. It is up to the individual domains how to ensure the QoS guarantee.

- QoS Reporting mechanisms may be activated to monitor the QoS provided in each domain to ensure the end-to-end QoS is provided at the end of the service instance.

Types of requirement

The end-to-end QoS requirements may include one or more of the following:

- Bandwidth.

- Latency.

- Jitter.

5.5 Charging

1) RACS should collect and provide charging information related to the resource reservation request, the resource modification request, the release and the abort request of the network resources.

2) RACS should collect and provide charging information related to the modification of QoS, during an active session in the intra domain and in the inter domain scenarios.

NOTE: The use of RACS QoS based charging information is an operator choice.

Annex A (informative): Scenarios

A.1 Use of sophisticated admission control algorithms

Admission control schemes can be classified into deterministic schemes and statistical schemes according to the way RACS handles the QoS requirements within a request sent by an AF.

Deterministic admission control schemes are characterized by worst-case assumptions, that is overbooking is prevented and reservations are based on the peak bandwidth. This approach provides a deterministic QoS to all users, but needlessly constrain the number of requests that can be granted simultaneously, and lead to an under-utilization of resources related to how bursty and variable the admitted traffic is.

Statistical schemes provide a probabilistic QoS guarantee instead of deterministic one since they reserve resources based on a probability distribution of the bandwidth consumption of VBR sources, resulting in higher resource utilization due to statistical multiplexing. The risk with such approach is that bearer flows can receive downgraded quality.

In the RACS you probably can implement a deterministic scheme till you manage CBR sources (e.g. voice traffic), although codec negotiation may imply under utilization of the resources since the use of the preferred codec within an offer is not mandatory. As soon as RACS manages variable bit-rate flows the waste of bandwidth caused by a deterministic approach of course increases. In that situation the RACS should probably implement a statistical method: the amount of bandwidth that a particular flow will need over a specific layer 2 technology can be determine based on a acceptable level of multiplexing and combining the peak requested bandwidth with the traffic characteristic of the flow (e.g. burstiness and packet size) RACS can be aware of for example thanks to the Reservation Class parameter in the Gq' interface.

In figure A.1 four different variable bit-rate flows are shown with the same peak bandwidth but with different characteristic as regard burstiness and average bandwidth. Using a deterministic approach based on the peak bandwidth for example only two requests could be granted. Using the average bandwidth as parameter all the flows should be admitted.

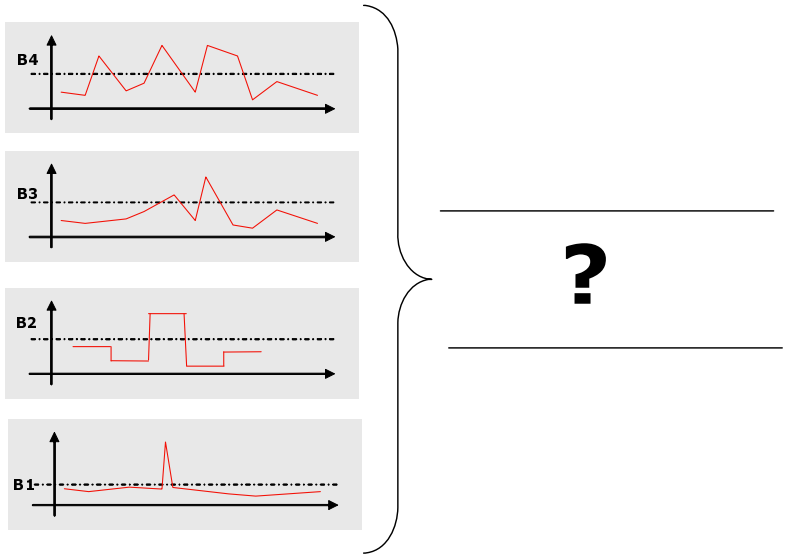


Figure A.1: Admission control for VBR flows

Having feedbacks of the quality bearer flows are achieving can be useful for improving RACS operating. Based on QoS reports you can, for example:

- tune the threshold (e.g. overbooking factor) within the admission algorithm. Referring to figure A.1, whether two flows have been already admitted and a third request is received it could be possible to accept the third request since for the admitted flows you receive a quality report well above a warning threshold. If then the new feedback RACS receives shows that the quality is lowering it could be possible to deny the four request although, for example, looking only at the average bandwidth all flows should be admitted;
- cope with variable bit-rate sources that behave differently from what is modelled within the RACS;
- better deal with priority and pre-emption mechanism. For example you can decide not to pre-empt a request if the quality of the already admitted flows is over a certain threshold although you have to admit a flow with an higher priority; or you can decide to perform the pre-emption operations only when a decrease in the quality of the bearer flows is notified.

A.2 L2 topology awareness and traffic management options

Let is consider a RACS controlling an ATM-based access network. In order to effective perform resource management, it should be aware of the L2 topology and of the traffic management options such as congestion control mechanisms implemented. In fact such mechanisms impact on the quality of the bearer flows received no matter which traffic conditioning policies you enforce both on the IP edge an access node.

Focusing on Connection Admission Control (CAC) mechanism within an ATM network different approaches can be used. The simplest form among all CAC algorithms, is the so-called Peak Bandwidth Allocation that uses only the knowledge of the PCR parameter to compare against the network available bandwidth and decide whether to accept the configuration of new connection or not. This algorithm ensures that the sum of requested resources and existing connections is bounded by the physical link capacity. but prevents any multiplexing gain among the VC and VP configured into the network. Another approach is to admit new connection allocating a bandwidth between the peak cell rate and the sustained cell rate. As a result, the sum of all the admitted connections' peak cell rates may be greater than the outgoing link capacity.

If a statistical approach is used in the ATM network, RACS should be aware of it and should exactly know for example the different overbooking factors used in all the interfaces of the various ATM switches. Referring to figure A.2 in order to admit a flow from UE-1 overbooking along the path towards the IP Edge should be taken into account.

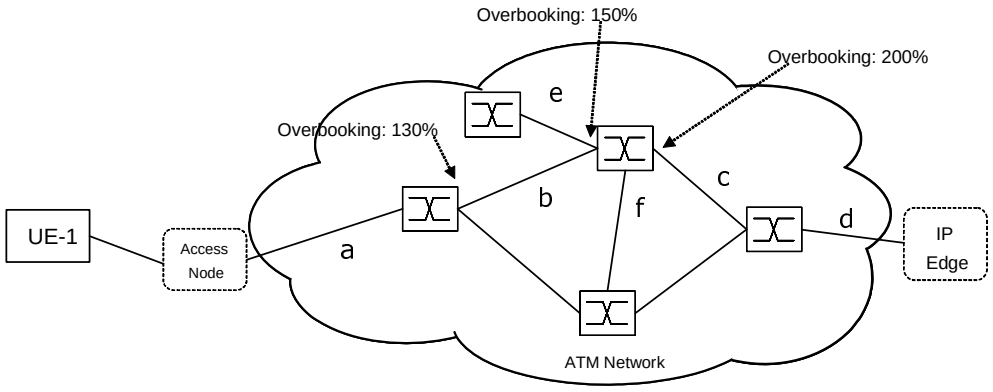


Figure A.2: L2 ATM Access Network

In such a scenario, if RACS can not collect all the needed information QoS reporting mechanism could help RACS to take into account the state within the ATM network without the need of gathering a lot of data and nevertheless being able to cope with all the situations. As soon as QoS reports degrade, RACS shall for example lower the available bandwidth.

A.3 Coexistence of managed and un-managed traffic

RACS receives reservation requests only for a subset of the traffic entering the network. Situations exist where the presence of un-managed traffic may impact on the performance of the higher priority flows managed by the RACS.

Let is consider the ATM access network depicted in figure A.2: over all the links (a, b, c,…) un-managed traffic can be present participating in the creation of congestion situation in which the managed traffic can be discarded although it should not according to the information available at the RACS level.

A similar situation may be foreseen considering queuing and scheduling configurations in the IP Edge node. In figure A.3, a hierarchical scheduler for managing traffic entering an ATM-based access network where Virtual Circuits are multiplexed within Virtual Path is shown. The number of queues varies from one level to another: a priority queue dedicated for Voice traffic is configured at the VC level but not at the VP level. In such a scenario, if a subscriber starts a voice call, the data traffic that he sends into the network impacts on the delay and on the bandwidth available for his voice traffic. Moreover also the data traffic of other users on the same VP impacts on his voice quality. As a final remark, also the data traffic of users on other VPs on the same BRAS port can impact voice quality whether a congestion situation in the ATM network occurs and data cells have not the CLP bit set to 0.

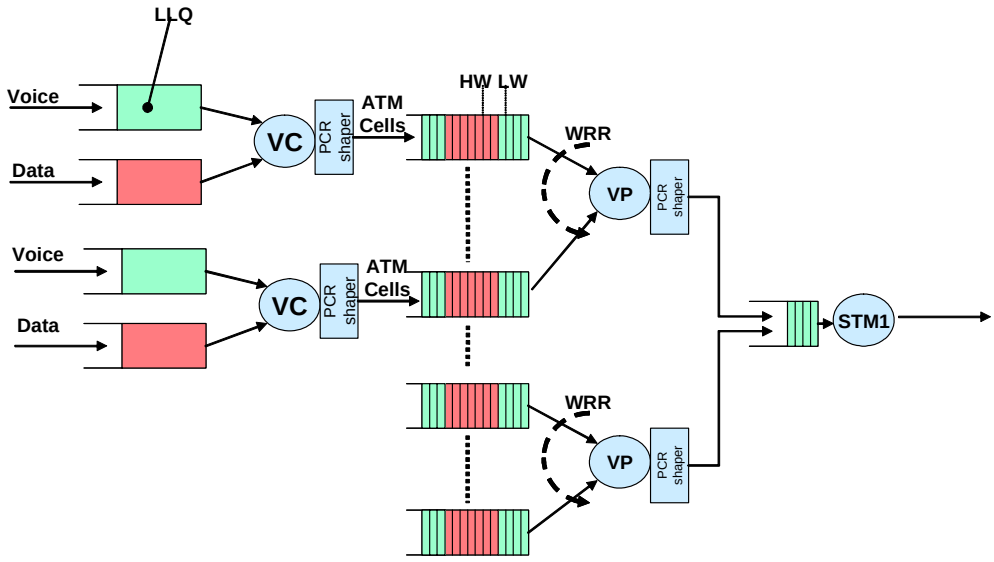


Figure 3: BRAS hierarchical scheduler

According to this scenario, QoS Reporting mechanism could help RACS to be informed of the un-managed traffic entering the network under its control and accordingly react. As soon as QoS reports degrade, RACS shall for example lower the available bandwidth.

A.4 Bandwidth on Demand

This scenario describes a Bandwidth on Demand (BoD) service offered to a User which allows the user to boost his access bandwidth to a higher level for a limited period of time.

The objective is to study the requirements on the RACS functionalities to deliver such a service. Note that charging aspects are considered out of scope for this example.

The service would allow users who usually have e.g. 512 Kbps of access bandwidth available to increase this to a higher bandwidth, e.g. 2 Mbps, for a limited period of time.

The service could offer a number of options:

1) Request for increased downstream bandwidth for a specified time. This would allow the user to e.g. download content at a higher rate than normal. In this case only downstream traffic would need to be boosted.

2) Request for increased bi-directional bandwidth for a specified time. This would allow the user to e.g. share content or take part in a video conference and enjoy an improved image and sound quality.

3) Request for increased bandwidth for a specific service which would only affect the traffic associated with that service, e.g. a "Video Boost" service.

4) Request for increased bandwidth for some specific content.

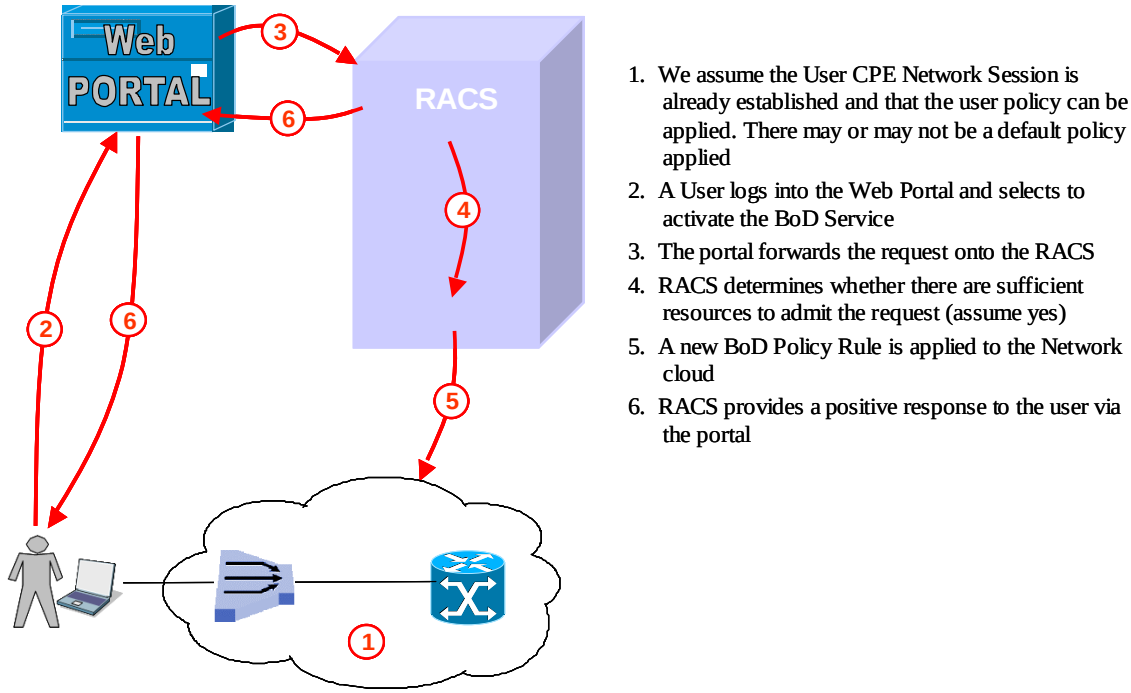
Of course any of the options above (time based, service based and content based) could be combined.

The service would benefit the user by allowing him to enjoy an improved QoS for a limited duration without the cost of subscribing to a permanent increase in bandwidth. The benefit to the Service Provider would be to offer a chargeable service to the user with the added incentive that the user may decide to upgrade to a higher bandwidth service on a permanent basis.

The traffic affected by a bandwidth change may be all of the traffic to/from a particular subscriber or only a specific subset related to the given service requested. For example, in option (3) above only the video traffic would benefit from the newly available bandwidth. Also, depending on the service requested, bandwidth changes may have an impact on upstream and/or downstream traffic.

Bandwidth rates and the services to which they apply may be explicitly requested by the user or the user may request pre-defined policies. The request could be performed through a web portal. When explicitly requesting a bandwidth rate, the user must include sufficient information to identify the traffic that should be boosted. When requesting a pre-defined policy, the user must minimally indicate the policy s/he wants to activate. In this last scenario the user/web portal may be unaware of the specific bandwidth change that RACS will apply in the network since RACS will hide this information.

Figure A.4 represents an example of a possible realization of this service.



1. We assume the User CPE Network Session is already established and that the user policy can be applied. There may or may not be a default policy applied
2. A User logs into the Web Portal and selects to activate the BoD Service
3. The portal forwards the request onto the RACS
4. RACS determines whether there are sufficient resources to admit the request (assume yes)
5. A new BoD Policy Rule is applied to the Network cloud
6. RACS provides a positive response to the user via the portal

Figure 4: Bandwidth-on-Demand scenario

The sequence of operations is as follows:

- The user logs on to a web portal and selects the BoD service.
- The Web Portal forwards the request to the RACS including an indication of quantity for bandwidth boost (this may be explicitly indicated in the request, or it is possible to reference pre-defined BoD classes).
- The RACS may determine whether there are sufficient network bandwidth resources to admit the request (we assume there are).

- A new policy is applied in the network for the subscriber according to the quantity of bandwidth requested.
- A positive response is forwarded to the user.
- If the Admission Control fails, the user is notified, and no upgrade is configured.

Annex B (informative): Bibliography

- ETSI TS 185 001: "Telecommunication and Internet converged Services and Protocols for Advanced Networking (TISPAN); Next Generation Network (NGN); Quality of Service (QoS) Framework and Requirements".
- ETSI ES 282 003: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); Resource and Admission Control Sub-system (RACS); Functional Architecture".
- ETSI TS 182 019: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); Resource and Admission Control Sub-system (RACS); Functional Architecture; Release 2".
- ETSI TR 182 022: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); Architectures for QoS handling".
- 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".
- ETSI TR 180 000: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); NGN Terminology".

History

Document history		
V2.0.0	August 2007	Publication