NSIS Working Group
Internet Draft M. Brunner (Editor)
Document: draft-ietf-nsis-req-02.txt NEC
Expires: November 2002 May 2002

Requirements for QoS Signaling Protocols
<draft-ietf-nsis-req-02.txt>

Status of this Memo

Abstract

This document defines requirements for signaling QoS across different network environments. To achieve wide applicability of the requirements, the starting point is a diverse set of scenarios/use cases concerning various types of networks and application interactions. We also provide an outline structure for the problem, including QoS related terminology. Taken with the scenarios, this allows us to focus more precisely on which parts of the overall QoS problem needs to be solved. We present the assumptions and the aspects not considered within scope before listing the requirements grouped according to areas such as architecture and design goals, signaling flows, layering, performance, flexibility, security, and mobility.

1 Introduction

This document defines requirements for signaling QoS across different network environments. It does not list any problems of existing QoS signaling protocols such as RSVP.

In order to derive requirements for QoS signaling it is necessary to first have a clear idea of the scope within which they are applicable.
We describe a set of QoS signaling scenarios and use cases in the Appendix of that document. These scenarios derive from a variety of backgrounds, and help obtain a clearer picture of what is in or out of scope of the NSIS work. They illustrate the problem of QoS signaling from various perspectives (end-system, access network, core network) and for various areas (fixed line, mobile, wireless environments). As the NSIS work becomes more clearly defined, scenarios will be added or dropped, or defined in more detail.

Based on these scenarios, we are able to define the QoS signaling problem on a more abstract level. In Section 3, we thus present a simple conceptual model of the QoS signaling problem, describe the entities involved in QoS signaling, and typical signaling paths. In Section 4 we list assumptions and exclusions.

The model of Section 3 allows deriving requirements from the scenarios presented in the appendix in a coherent and consistent manner. Requirements are grouped according to areas such as Architecture and design goals, Signaling Flows, Layering, Performance, Flexibility, Security and Mobility.

QoS is a pretty large field with a lot of interaction with other protocols, mechanisms, applications etc. In the following, some thoughts from an end-system point of view and from a network point of view.

End-system perspective: In future mobile terminals, the support of adaptive applications is more and more important. Adaptively can be seen as an important technique to react to QoS violations that may occur frequently, e.g., in wireless environments due to changed environmental and network conditions. This may result in degraded end-to-end performance. It is then up to adaptive applications to react to the new resource availability. Therefore, it is essential to define interoperability between media-, mobility- and QoS management. While most likely mobile terminals cannot assume, that explicit QoS reservation schemes are available, some access networks nevertheless may offer such capabilities. Applications subscribed to an end-system QoS management system should be supported with a dedicated QoS API to set-up, control and adapt media sessions.

Network perspective: QoS enabled IP networks are expected to handle two different kinds of QoS granularities: per-flow QoS and per-trunk/per-class QoS. Per-flow QoS might be needed in access networks and may there be subject of QoS signaling. However, in the core network only per-trunk or per-class QoS can be considered for scalability reasons. Therefore there might be different requirements on QoS signaling applying to different parts of the network. In the access network QoS signaling is an interaction between end systems and access routers or access network QoS managers (in the following we call them QoS initiator and QoS controller). In the core network QoS signaling refers to trunks or classes of traffic between core and edge systems or between peering core systems. Please note that this does not exclude the transport of per-flow signaling through core networks.

It is clear from these descriptions that the subject of QoS is uniquely complex and any investigation could potentially have a very broad scope - so broad that it is a challenge to focus work on an area which could lead to a concrete and useful result. This is our motivation for considering a set of use cases, which map out the domain of application that we want to address. It is also the motivation for defining a problem structure, which allows us to state the boundaries of what types of functionality to consider, and to list background assumptions.

There are several areas of the requirements related to networking aspects which are incomplete, for example, interaction with host and site multi-homing, use of anycast services, and so on. These issues should be considered in any future requirement analysis work.

2 Terminology

In the area of Qualiaty of Service (QoS) it is quite difficult and an exercise for its own to define terminology. Nevertheless, we tried to list the most often used terms in the draft and tried to explain them. However, don't be to religious about it, they are not meant to prescribe any thing in the draft.

Aggregate: a group of flows, usually with similar QoS requirements, which can be treated together as a whole with a single overall QoS requirement for signaling and provisioning. Aggregates and flows can be further aggregated together.

[QoS] Domain: a collection of networks under the same administrative control and grouped together for administrative purposes.

Egress point: the router via which a path exits a domain/subdomain.

End Host: the end system or host, for whose flows QoS is being requested and provisioned.

End-to-End QoS: the QoS delivered by the network between two communicating end hosts. End-to-end QoS co-ordinates and enforces predefined traffic management policies across multiple network entities and administrative domains.

Edge-to-edge QoS: QoS within an administrative domain that connects to other networks rather than hosts or end systems.

Flow: a traffic stream (sequence of IP packets between two end systems) for which a specific level of QoS is to be provided. The flow can be unicast (uni- or bi-directional) or multicast.

Flow Administration: represents the policy associated with how flows should be treated in the network, for example whether and how the flows should be aggregated. It may consist of both user and local network management information.

Higher Layers: the higher layer (transport protocol and application) functions that request QoS from the network layer. The request might be a trigger generated within the end system, or the trigger might be provided by some entity within the network (e.g. application proxy or policy server).

Indication: feedback from QoS provisioning to indicate the current QoS being provided to a flow or aggregate, and whether any violations have been detected by the QoS technology being used within the local domain/subdomain.

Ingress point: the router via which a path enters a domain/subdomain.

Mapping: the act of transforming parameters from QSCs to values that are meaningful to the actual QoS technology in use in the domain/subdomain.

Path: the route across the networks taken by a flow or aggregate, i.e. which domains/subdomains it passes through and the egress/ingress points for each.

Path segment: the segment of a path within a single domain/subdomain.

QoS Administration Function: a generic term for all functions associated with admission control, policy control, traffic engineering etc.

QoS Control Information: the information the governs the QoS treatment to be applied to a flow or aggregate, including the QSC, flow administration, and any associated security or accounting information.

QoS Controller: this is responsible for interpreting the signaling carrying the user QoS parameters, optionally inserting/modifying the parameters according to local network QoS management policy, and invoking local QoS provisioning mechanisms. Note that q QoS controller might have very different functionality depending on where in the network and in what environment they are implemented.

QoS Initiator: this is responsible for generating the QSCs for traffic flow(s) based on user or application requirements and signaling them to the network as well as invoking local QoS provisioning mechanisms. This can be located in the end system, but may reside elsewhere in network.

QoS Provisioning: the act of actually allocating resources to a flow or aggregate of flows, may include mechanisms such as LSP initiation for MPLS, packet scheduler configuration within a router, and so on. The mechanisms depend on the overall QoS technology being used within the [sub]domain.

QoS Service Classes (QSC): specify the QoS requirements of a traffic flow or aggregate. Can be further sub-divided into user specific and network related parameters

QoS Signaling: a way to communicate QSCs and QoS management information between hosts, end systems and network devices etc. May include request and response messages to facilitate negotiation/re-negotiation, asynchronous feedback messages (not delivered upon request) to inform End Hosts, QoS initiators and QoS controllers about current QoS levels, and QoS querying facilities.

[QoS] Subdomain: a network within an administrative domain using a uniform technology/QoS provisioning function to provision resources.

QoS Technology: a generic term for a set of protocols, standards and mechanisms that can be used within a QoS domain/subdomain to manage the QoS provided to flows or aggregates that traverse the domain. Examples might include MPLS, DiffServ, and so on. A QoS technology is associated with certain QoS provisioning techniques.

QoS Violation: occurs when the QoS applied to a flow or aggregate does not meet the requested and negotiated QoS agreed for it.

Resource: something of value in a network infrastructure to which rules or policy criteria are first applied before access is granted. Examples of resources include the buffers in a router and bandwidth on an interface.

Resource Allocation: part of a resource that has been dedicated for the use of a particular traffic type for a period of time through the application of policies.

Sender-initiated QoS signaling protocol: A sender-initiated QoS signaling protocol is a protocol (see e.g., YESSIR [8], RMD [10]) where the QI initiates the signaling on behalf of the sender of the data. What this means is that admission control and resource allocation functions are processed from the data sender towards the data receiver. However, the triggering instance is not specified.

Receiver-initiated QoS signalling protocol: A receiver-initiated protocol, (see e.g., RSVP [9]) is a protocol where the QoS reservations are initiated by the QoS Reiceiver on behalf of the receiver of the user data. What this means is that admission control and resource allocation functions are processed from the data receiver back towards the data sender. However, the triggering instance is not specified.

3 Problem Statement and Scope

We provide in the following a preliminary architectural picture as a basis for discussion. We will refer to it in the following requirement section.

A set of issues and problems to be solved has been given at a top level by the use cases/scenarios of the appendix. However, the problem of QoS has an extremely wide scope and there is a great deal of work already done to provide different components of the solution, such as QoS technologies for example. A basic goal should be to re-use these wherever possible, and to focus requirements work at an early stage on those areas where a new solution is needed (e.g. an especially simple one). We also try to avoid defining requirements related to internal implementation aspects.

In this section, we present a simple conceptual model of the overall QoS problem in order to identify the applicability to NSIS of requirements derived from the use cases, and to clarify the scope of the work, including any open issues. This model also identifies further sources of requirements from external interactions with other parts of an overall QoS solution, clarifies the terminology used, and allows the statement of design goals about the nature of the solution (see section 5).

Note that this model is intended not to constrain the technical approach taken subsequently, simply to allow concrete phrasing of requirements (e.g. requirements about placement of the QoS initiator, or ability to 'drive' particular QoS technologies.)

Roughly, the scope of NSIS is assumed to be the interaction between the QoS initiator and QoS controller(s), including selection of signaling protocols to carry the QoS information, and the syntax/semantics of the information that is exchanged. Further statements on assumptions/exclusions are given in the next Section. The main elements are:

1. Something that starts the request for QoS, the QoS Initiator.

This might be in the end system or within some other part of the network. The distinguishing feature of the QoS initiator is that it acts on triggers coming (directly or indirectly) from the higher layers in the end systems. It needs to map the QoS requested by them, and also provides feedback information to the higher layers which might be used by transport layer rate management or adaptive applications.

2. Something that assists in managing QoS further along the path, the QoS controller.

The QoS controller does not interact with higher layers, but interacts with the QoS initiator and possibly more QoS controllers on the path, edge to edge or possibly end to end.

3. The QoS initiator and controller(s) interact with each other, path segment by path segment. This interaction involves the exchange of data (QoS control information) over some signaling protocol.

4. The path segment traverses an underlying network (QoS domain or subdomain) covering one or more IP hops. The underlying network uses some local QoS technology. This QoS technology has to be provisioned appropriately for the flow, and this is done by the QoS initiator and controller(s), mapping their QoS control information to technology-related QoS parameters and receiving indications about success or failure in response.

Now concentrating more on the overall end to end (multiple QoS domains) aspects, in particular:

1. The QoS initiator need not be located at an end system, and the QoS controllers are not assumed to be located on the flow's data path. However, they must be able to identify the ingress and egress points for the flow path as it traverses the domain/subdomain. Any signaling protocol must be able to find the appropriate QoS controller and carry this ingress/egress point information.

2. We see the network at the level of domains/subdomains rather than individual routers (except in the special case that the domain contains one link). Domains are assumed to be administrative entities, so security requirements apply to the signaling between them. Subdomains are introduced to allow the fact a given QoS provisioning mechanism may only be used within a part of a domain, typically for a particular subnetwork technology boundary. Aggregation can also take place at subdomain boundaries.

3. Any domain may contain QoS administration functions (e.g. to do with traffic engineering, admission control, policy and so on).

These are assumed to interact with the QoS initiator and controllers (and end systems) using standard mechanisms.

4. The placement of the QoS initiators and QoS controllers is not fixed. Actually, there are two extreme cases:

- Each router on the data path implements a QoS controller and QoS initiator.

- Only the end systems incorporate a QoS controller and QoS initiator, which means the end systems need to have QoS provisioning capabilities. However this case does not seam to be realistic but shows the flexible allocation of the controller and initiator function.

4 Assumptions and Exclusions

4.1 Assumptions and Non-Assumptions

1. The NSIS signaling could run end to end, end to edge, or edge to edge, or network-to-network ((between providers), depending on what point in the network acts as the initiator, and how far towards the other end of the network the signaling propagates. Although the figures show QoS controllers at a very limited number of locations in the network (e.g. at domain or subdomain borders, or even controlling a complete domain), this is only one possible case. In general, we could expect QoS controllers to become more 'dense' towards the edges of the network, but this is not a requirement. An overprovisioned domain might contain no QoS controllers at all (and be NSIS transparent); at the other extreme, QoS controllers might be placed at every router. In the latter case, QoS provisioning can be carried out in a local implementation-dependent way without further signalling, whereas in the case of remote QoS controllers, a provisioning protocol might be needed to control the routers along the path. This provisioning protocol is then independent of the end to end NSIS signalling.

2. We do not consider 'pure' end-to-end QoS signaling that is not interpreted anywhere within the network. Such signaling is an application-layer issue and IETF protocols such as SIP etc. can be used.

3. Where the signaling does cover several QoS domains or subdomains, we do not exclude that different signaling protocols are used in each path segment. We only place requirements on the universality of the QoS control information that is being transported. (The goals here would be to allow the use of signaling protocols which are matched to the characteristics of the portion of the network being traversed.) Note that the outcome of NSIS work might result in

various protocols or various flavors of the same protocol. This implies the need for the translation of information into QoS domain specific format as well.

4. We assume that the service definitions a QoS initiator can ask for are known in advance of the signaling protocol running. Service definition includes QoS parameters, life-time of QoS guarantee etc. There are many ways a service requester get to know about it. There might be standardized services, the definition can be negotiated together with a contract, the service definition is published at a Web-page, etc.

5. We assume that there are means for the discovery of NSIS entities in order to know the signaling peers (solutions include static configuration, automatically discovered, or implicitly runs over the right nodes, etc.)

4.2 Exclusions

1. Development of specific mechanisms and algorithms for application and transport layer adaptation are not considered, nor are the protocols that would support it.

2. Specific mechanisms (APIs and so on) for interaction between transport/applications and the network layer are not considered, except to clarify the requirements on the negotiation capabilities and information semantics that would be needed of the signaling protocol. The same applies to application adaptation mechanisms.

3. Specific mechanisms for QoS provisioning within a domain/subdomain are not considered. It should be possible to exploit these mechanisms optimally within the end to end context. Consideration of how to do this might generate new requirements for NSIS however. For example, the information needed by an QoS controller to manage a radio subnetwork needs to be provided by the NSIS solution.

4. Specific mechanisms (APIs and so on) for interaction between the network layer and underlying QoS provisioning mechanisms are not considered.

5. Interaction with QoS administration capabilities is not considered. Standard protocols should be used for this (e.g. COPS). This may imply requirements for the sort of information that should be exchanged between the NSIS network QoS entities.

6. Security issues related to multicasting are outside the scope of the QoS signaling protocol.

Since multicasting is currently not an issue for the QoS protocol, security issues related to multicast are outside the scope. Multicast security may additionally be an application issue that is also outside the scope of the QoS protocol.

7. Protection of non-QoS signaling messages is outside the scope of the QoS protocol

Security protection of data messages transmitted along the established QoS path is outside the scope of the QoS protocol. These security properties are likely to be application specific and may be provided by the corresponding application layer protocol.

8. Service definitions and QoS classes are out of scope. Together with the service definition any definition of service specific parameters are not considered in this draft. Only the base NSIS signaling protocol for transporting the QoS/service information are handled.

9. Similarly, specific methods, protocols, and ways to express QoS information in the Application/Session level are not considered (e.g., SDP, SIP, RTSP, etc.).

10. The specification of any extensions needed to signal QoS information via application level protocols (e.g. SDP(ng)), and the mapping on NSIS information are considered outside of the scope of NSIS working group, as this work is in the direct scope of other IETF working groups (e.g. MMUSIC).

5 Requirements

This section defines more detailed requirements for a QoS signaling solution, derived from consideration of the use cases/scenarios, and respecting the framework, scoping assumptions, and terminology considered earlier. The requirements are in subsections, grouped roughly according to general technical aspects: architecture and design goals, topology issues, QoS parameters, performance, security, information, and flexibility.

Two general (and potentially contradictory) goals for the solution are that it should be applicable in a very wide range of scenarios, and at the same time lightweight in implementation complexity and resource requirements in nodes. One approach to this is that the solution could deal with certain requirements via modular components or capabilities, which are optional to implement in individual nodes.

Some of the requirements are technically contradictory. Depending on the scenarios a solution applies to, one or the other requirement is applicable.

Find in Section 6 the MUSTs, SHOULDs, and MAYs

5.1 Architecture and Design Goals

This section contains requirements related to desirable overall characteristics of a solution, e.g. enabling flexibility, or independence of parts of the framework.

5.1.1 Applicability for different QoS technologies.

The QoS signaling protocol must work with various QoS technologies. The information exchanged over the signaling protocol must be in such detail and quantity that it is useful for various QoS technologies.

5.1.2 Resource availability information on request

In some scenarios, e.g., the mobile terminal scenario, it is required to query, whether resources are available, without performing a reservation on the resource. One solution might be a feedback mechanism based on which a QoS inferred handover can take place.

5.1.3 Modularity

A modular design allows for more lightweight implementations, if fewer features are needed. Mutually exclusive solutions are supported. Examples for modularity:

- Work over any kind of network (narrowband / broadband, error-prone / reliable...) - This implies low bandwidth signaling and redundant information must be supported if necessary.

- In case QoS requirements are soft (e.g. banking transactions, gaming), fast and lightweight signaling (e.g., not more than one round-trip time)

- Uni- and bi-directional reservations are possible

5.1.4 Decoupling of protocol and information it is carrying

The signaling protocol(s) used must be clearly separated from the QoS control information being transported. This provides for the independent development of these two aspects of the solution, and allows for this control information to be carried within other protocols, including application layer ones, existing ones or those being developed in the future. The gained flexibility in the information transported allows for the applicability of the same protocol in various scenarios.
However, note that the information carried needs to be the same. Otherwise interoperability is difficult to achieve.

5.1.5 Reuse of existing QoS provisioning

Reuse existing QoS functions and protocols for QoS provisioning within a domain/subdomain unchanged. (Motivation: 'Don't re-invent the wheel'.)

5.1.6 Independence of signaling and provisioning paradigm

The QoS signaling should be independent of the paradigm and mechanism of QoS provisioning. The independence allows for using the NSIS protocol together with various QoS technologies.

5.2 Signaling Flows

This section contains requirements related to the possible signaling flows that should be supported, e.g. over what parts of the flow path, between what entities (end-systems, routers, middleboxes, management systems), in which direction.

5.2.1 Free placement of QoS Initiator and QoS Controllers functions

The protocol(s) must work in various scenarios such as host-to-network-to-host, edge-to-edge, (e.g., just within one providers domain), user-to-network (from end system into the network, ending, e.g., at the entry to the network and vice versa), network-to-network (e.g., between providers).

Placing the QoS controller and initiator functions at different locations allows for various scenarios to work with the same or similar protocols.

5.2.2 No constraint of the QoS signaling and QoS Controllers to be in the data path.

There is a set of scenarios, where QoS signaling is not on the data path. The QoS Controller being in the data path is one extreme case and useful in certain cases.

There are going to be cases where a centralized entity will take a decision about QoS requests. In this case, there's no question there is no need to have data follow the signalling path.

There are going to be cases wiout a centralized entity managing resources and the signaling will be used as a tool for resource management. For various reasons (such as efficient use of expensive bandwidth), one will want to have fine-grained, fast, and very dynamic control of the resources in the network. -

There are going to be cases where there will be neither signaling nor a centralized entity (overprovisioning). Nothing has to be done anyway.

One can capture the requirement with the following wording: If one views the domain with a QoS technology as a virtual router then NSIS signaling used between those virtual routers must follow the same path as the data.

Routing the signaling protocol along an independent path is desired by network operators/designers. Ideally, the capability to route the protocol along an independent path would give the network designer/operator the option to manage bandwidth utilization through the topology.

There are other possibilities as well. An NSIS protocol must accept all of these possibilities.

5.2.3 Concealment of topology and technology information

The QoS protocol should allow hiding the internal structure of a QoS domain from end-nodes and from other networks. Hence an adversary should not be able to learn the internal structure of a network with the help of the QoS protocol.

In various scenarios, topology information should be hidden for various reasons. From a business point of view, some administrations don't want to reveal the topology and technology used.

5.2.4 Optional transparency of QoS signaling to network

It should be possible that the QoS signaling for some flows traverse path segments transparently, i.e., without interpretation at QoS controllers within the network. An example would be a subdomain within a core network, which only interpreted signaling for aggregates established at the domain edge, with the flow-related signaling passing transparently through it.

5.3 Additional information beyond signaling of QoS information

This section contains the desired signaling (messages) for other purposes other than that for conveying QoS parameters.

5.3.1 Explicit release of resources

When a QoS reservation is no longer necessary, e.g. because the application terminates, or because a mobile host experienced a hand-off, it must be possible to explicitly release resources.

5.3.2 Possibility for automatic release of resources after failure

When the QoS Initiator goes down, the resources it requested in the network should be released, since they will no longer be necessary.

After detection of a failure in the network, any QoS controller/initiator must be able to release a reservation it is involved in. For example, this may require signaling of the "Release after Failure" message upstream as well as downstream, or soft state timing out of reservations.

Note that this might need to work together with a notification mechanism.

5.3.3 Possibility for automatic re-setup of resources after recovery

In case of a failure, the reservation can get setup again automatically. It enables sort of a persistent reservation, if the QoS Initiator requests it. In scenarios where the reservations are on a longer time scale, this could make sense to reduce the signaling load in case of failure and recovery.

5.3.4 Prompt notification of QoS violation in case of error/failure to QoS Initiator and QoS Controllers

QoS Controllers should be able to notify the QoS Initiator, if there is an error inside the network. There are two types of network errors:

Recoverable errors: This type error can be locally repaired by the network nodes. The network nodes do not have to notify the users of the error immediately. This is a condition when the danger of degradation (or actual short term degradation) of the provided QoS was overcome by the network (QoS controller) itself.

Unrecoverable errors: the network nodes cannot handle this type of error, and have to notify the users as soon as possible.

5.3.5 Feedback about success of request for QoS guarantees

A request for QoS must be answered at least with yes or no. However, it might be useful in case of a negative answer to also get a description of what might be the QoS one can successfully request etc. So it might be useful to include an opaque element into the answer. The element heavily depends on the service requested.

5.3.6 Allow local QoS information exchange between nodes of the same administrative domain

The QoS signaling protocol must be able to exchange local QoS information between QoS controllers located within one single domain. Local QoS information might, for example, be IP addresses, severe congestion notification, notification of successful or erroneous processing of QoS signaling messages.

In some cases, the NSIS QoS signalling protocol may carry identification of the QoS controllers located at the boundaries of a domain. However, the identification of edge should not be visible to the end host (QoS initiator) and only applies within one QoS administrative domain.

5.4 Layering

This section contains requirements related to the way the signaling being considered interacts with upper layer functions (users, applications, and QoS administration), and lower layer QoS technologies.

5.4.1 The signaling protocol and QoS control information should be application independent.

However, opaque application information might get transported in the signaling message, without being handled in the network. Development and deployment of new applications should be possible without impacting the network infrastructure. Additionally, QoS protocols are expected to conform to the Internet principles.

5.5 QoS Control Information

This section contains requirements related to the QoS control information that needs to be exchanged.

5.5.1 Mutability information on parameters

It should be possible for the initiator to control the mutability of the QSC information. This prevents from being changed in a non-recoverable way. The initiator should be able to control what is requested end to end, without the request being gradually mutated as it passes through a sequence of domains. This implies that in case of changes made on the parameters, the original requested ones must still be available.

Note that we do not require anything about particular QoS paramters being changed.

5.5.2 Possibility to add and remove local domain information

It should be possible for the QoS control functions to add and remove local scope elements. E.g., at the entrance to a QoS domain domain-specific information is added, which is used in this domain only, and the information is removed again when a signaling message leaves the domain. The motivation is in the economy of re-use the protocol for domain internal signaling of various information. Where additional information is needed for QoS control within a particular domain, it should be possible to carry this at the same time as the 'end to end' information.)

5.5.3 Independence of reservation identifier

A reservation identifier must be used, which is independent of the flow identifier, the IP address of the QoS Initiator, and the flow end-points. Various scenarios in the mobility area require this independence because flows resulting from handoff might have changed end-points etc. but still have the same QoS requirement.

5.5.4 Seamless modification of already reserved QoS

In many case, the reservation needs to be updated (up or downgrade). This must happen seamlessly without service interruption. At least the signaling protocol must allow for it, even if some data path elements might not be capable of doing so.

5.5.5 Signaling must support quantitative, qualitative, and relative QoS specifications

5.6 Performance

This section discusses performance requirements and evaluation criteria and the way in which these could and should be traded off against each other in various parts of the solution.

Scalability is a must anyway. However, depending on the scenario the question to which extends the protocol must be scalable.

5.6.1 Scalability in the number of messages received by a signaling communication partner (QoS initiator and controller)

5.6.2 Scalability in number of hand-offs

5.6.3 Scalability in the number of interactions for setting up a reservation

5.6.4 Scalability in the number of state per entity (QoS initiators and QoS controllers)

5.6.5 Scalability in CPU use (end terminal and intermediate nodes)

5.6.6 Low latency in setup

Low latency is only needed in scenarios, where reservations are in a short time scale (e.g. handover in mobile environments), or where human interaction is immediately concerned (e.g., voice communication setup delay)

5.6.7 Allow for low bandwidth consumption for signaling protocol

Again only small sets of scenarios call for low bandwidth, mainly those where wireless links are involved.

Note that many of the performance issues are heavily dependent on the scenario assumed and are normally a trade-off between speed, reliability, complexity, and scalability. The trade-off varies in different parts of the network. For example, in radio access networks low bandwidth consumption will overweight the low latency requirement, while in core networks it may be reverse.

5.6.8 Ability to constrain load on devices

The NSIS architecture should give the ability to constrain the load (CPU load, memory space, signaling bandwidth consumption and signaling intensity) on devices where it is needed. One of the reasons is that the protocol handling should have a minimal impact on interior (core) nodes.

This can be achieved by many different methods. Examples, and this are only examples, include message aggregation, by ignoring signaling message, header compression, or minimizing functionality. The framework may choose any method as long as the requirement is met.

5.6.9 Highest possible network utilization

There are networking environments that require high network utilization for various reasons, and the signaling protocol should to its best ability support high resource utilization while maintaining appropriate QoS.

In networks where resources are very expensive (as is the case for many wireless networks), efficient network utilization is of critical financial importance. On the other hand there are other parts of the network where high utilization is not required.

5.7 Flexibility

This section lists the various ways the protocol can flexibly be employed.

5.7.1 Aggregation capability, including the capability to select and change the level of aggregation.

5.7.2 Flexibility in the placement of the QoS initiator

It might be the sender or the receiver of content. But also network-initiated reservations are required in various scenarios.

5.7.3 Flexibility in the initiation of re-negotiation (QoS change requests)

Again the sender or the receiver of content might initiate a re-negotiation due to various reasons, such as local resource shortage (CPU, memory on end-system) or a user changed application preference/profiles. But also network-initiated re-negotiation is required in cases, where the network is not able to further guarantee resources etc.

5.7.4 Uni / bi-directional reservation

Both uni-directonal as well as bi-direction reservations must be possible.

5.8 Security

This section discusses security-related requirements. First a list of security threats is given.

5.8.1 The QoS protocol must provide strong authentication

A QoS protocol must make provision for enabling various entities to be authenticated against each other using data origin and/or entity authentication. The QoS protocol must enable mutual authentication between the two communicating entities. The term strong authentication points to the fact that weak plain-text password mechanisms must not be used for authentication.

5.8.2 The QoS protocol must provide means to authorize resource requests

This requirement demands a hook to interact with a policy entity to request authorization data. This allows an authenticated entity to be associated with authorization data and to verify the resource request. Authorization prevents reservations by unauthorized entities, reservations violating policies, theft of service and additionally limits denial of service attacks against parts of the network or the entire network. Additionally it might be helpful to provide some means to inform other protocols of participating nodes within the same administrative domain about a previous successful authorization event.

5.8.3 The QoS signaling messages must provide integrity protection.

The integrity protection of the transmitted signaling messages prevent an adversary from modifying parts of the QoS signaling message and from mounting denial of service attacks against network elements participating in the QoS protocol.

5.8.4 The QoS signaling messages must be replay protected.

To prevent replay of previous signaling messages the QoS protocol must provide means to detect old messages. A solution must cover issues of synchronization problems in the case of a restart or a crash of a participating network element. The use of replay mechanism apart from sequence numbers should be investigated.

5.8.5 The QoS signaling protocol must allow for hop-by-hop security.

Hop-by-Hop security is a well known and proven concept in QoS protocols that allows intermediate nodes that actively participate in the QoS protocol to modify the messages as required by the QoS processing. Note that this requirement does not exclude end-to-end or network-to-network security of a QoS reservation request. End-to-end security between the initiator and the responder may be used to provide protection of non-mutable data fields. Network-to-network security refers to the protection of messages over various hops but not in an end-to-end manner i.e. protected over a particular network.

5.8.6 The QoS protocol should allow identity confidentiality and location privacy.

Identity confidentiality enables privacy and avoids profiling of entities by adversary eavesdropping the signaling traffic along the path. The identity used in the process of authentication may also be hidden to a limited extent from a network to which the initiator is attached. It is however required that the identity provide enough information for the access network to collect accounting data. Location privacy is an issue for the initiator who triggers the QoS protocol. In some scenarios the initiator may not be willing to reveal location information to the responder.

5.8.7 The QoS protocol should prevent denial-of-service attacks against signaling entities.

To effectively prevent denial-of-service attacks the QoS protocol and the used security mechanisms should not force to do heavy computation to verify a resource request prior authenticating the requesting entity. Additionally the QoS protocol and the used security mechanisms should not require large resource consumption (for example main memory or other additional message exchanges) before a successful authentication was done.

5.8.8 The QoS protocol should support confidentiality of signaling messages.

Based on the signaling information exchanged between nodes participating in the QoS protocol an adversary may learn both the identities and the content of the QoS messages. To prevent this from happening, confidentiality of the QoS requests in a hop-by-hop manner should be provided. Note that hop-by-hop is always required whenever entities actively participating in the protocol must be able to read and eventually modify the content of the QoS messages. This does not exclude the case where one or more network elements are not required to read the information of the transmitted QoS messages.

5.8.9 The QoS protocol should provide hooks to interact with protocols that allow the negotiation of authentication and key management protocols.

The negotiation of an authentication and key management protocols within the QoS protocol is outside the scope of the QoS protocol. This requirement originates from the fact that more than one key management protocol may be used to provide security associations. So both entities must be capable to use the same protocol which may be difficult in a mobile environment with different requirements and different protocols. The goal of such a negotiation step is to determine which authentication and key management protocol to use is executed prior to the execution of the chosen key management protocol. The used key management protocol must however be able to create a security association that matches with the one used in the QoS protocol. A QoS protocol should however provide a way to interact with these negotiation protocols.

5.8.10 The QoS protocol should provide means to interact with key management protocols

Key management protocols typically require a larger number of messages to be transmitted to allow a session key and the corresponding security association to be derived. To avoid the complex issue of mapping individual authentication and key management protocols to a QoS protocol such a protocol is outside the scope of the QoS protocol. Although the key management protocol may be independent there must be a way for the QoS protocol to exploit existing security associations to avoid executing a separate key management protocol (or instance of the same protocol) for protocols that closely operate together. If no such security association exists then there should be means for the QoS protocol to trigger a key management protocol to dynamically create the required security associations.

5.9 Mobility

5.9.1 Allow efficient QoS re-establishment after handover

Handover is an essential function in wireless networks. After handover, QoS may need to be completely or partially re-established due to route changes. The re-establishment may be requested by the mobile node itself or triggered by the access point that the mobile node is attached to. In the first case, the QoS signalling should allow efficient QoS re-establishment after handover. Re-establishment of QoS after handover should be as quick as possible so that the mobile node does not experience service interruption or QoS degradation. The re-establishment should be localized, and not require end-to-end signalling, if possible.

TBD

5.10 Interworking with other protocols and techniques

Hooks must be provided to enable efficient interworking between various protocols and techniques including:

5.10.1 Interworking with IP tunneling

IP tunneling for various applications must be supported. More specifically tunneling for IPSec tunnels are of importance. This mainly impacts the identification of flows. Additionally, care needs to be taken using IPSec for signaling message.

5.10.2 The solution should not constrain either to IPv4 or IPv6

5.10.3 Independence from charging model

Signaling must not be constrained by charging models or the charging infrastructure used. However, the end-system should be able to query current pay statistics and to specify user cost functions.

5.10.4 The QoS protocol should provide hooks for AAA protocols

The security mechanism should be developed with respect to be able to collect usage records from one or more network elements.

5.11 Operational

5.11.1 Ability to assign transport quality to signaling messages

The NSIS architecture should allow the network operator to assign the NSIS protocol messages a certain transport quality. As signaling opens up for possible denial-of-service attacks, this requirement gives the network operator a mean, but also the obligation, to trade-off between signaling latency and the impact (from the signaling messages) on devices within his/her network. From protocol design this requirement states that the protocol messages should be detectable, at least where the control and assignment of the messages priority is done.

6 The MUSTs, SHOULDs, and MAYs

In order to prioritize the various requirements from Section 5, we define different 'parts of the network'. In the different parts of the network a particular requirement might have a different priority.

The parts of the networks we differentiate are the host-to-first router, the access network, and the core network. The host to first router part includes all the layer 2 technologies to access to the Internet. In many cases, there is an application and/or user running on the host initiating QoS signaling. The access network can be characterized by low capacity links, meadium speed IP processing capabilities, and it might consist of a complete layer 2 network as well. The core network characteristics include high-speed forwarding capacities and interdomain QoS issues. All of them are not strictly defined and should not be regarded as that, but should give a feeling about where in the network we have different requirements concerning QoS signaling.

Note that the requirement titles are listed for better reading.

5.1 Architecture and Design Goals
5.1.1 Applicability for different QoS technologies.
5.1.2 Resource availability information on request
5.1.3 Modularity
5.1.4 Decoupling of protocol and information it is carrying
5.1.5 Reuse of existing QoS provisioning
5.1.6 Independence of signaling and provisioning paradigm

	host-to-net	access	core
5.1.1			

5.1.2				
5.1.3				

5.1.4				
5.1.5				
5.1.6				

5.2 Signaling Flows
5.2.1 Free placement of QoS Initiator and QoS Controllers functions

5.2.2 No constraint of the QoS signaling and QoS Controllers to be in the data path.
5.2.3 Concealment of topology and technology information
5.2.4 Optional transparency of QoS signaling to network

	host-to-net	access	core
5.2.1			
5.2.2			
5.2.3			
5.2.4			

5.3 Additional information beyond signaling of QoS information
5.3.1 Explicit release of resources
5.3.2 Possibility for automatic release of resources after failure
5.3.3 Possibility for automatic re-setup of resources after recovery
5.3.4 Prompt notification of QoS violation in case of error / failure to QoS Initiator and QoS Controllers
5.3.5 Feedback about success of request for QoS guarantees
5.3.6 Allow local QoS information exchange between nodes of the same administrative domain

	host-to-net	access	core
5.3.1			
5.3.2			
5.3.3			
5.3.4			

5.3.5				
5.3.6				

5.4 Layering

5.4.1 The signaling protocol and QoS control information should be application independent.

	host-to-net	access	core
5.4.1			

5.5 QoS Control Information

5.5.1 Mutability information on parameters
5.5.2 Possibility to add and remove local domain information
5.5.3 Independence of reservation identifier
5.5.4 Seamless modification of already reserved QoS
5.5.5 Signaling must support quantitative, qualitative, and relative QoS specifications

	host-to-net	access	core
5.5.1			
5.5.2			
5.5.3			
5.5.4			
5.5.5			

5.6 Performance

5.6.1 Scalability in the number of messages received by a signaling communication partner (QoS initiator and controller)
5.6.2 Scalability in number of hand-offs
5.6.3 Scalability in the number of interactions for setting up a reservation
5.6.4 Scalability in the number of state per entity (QoS initiators and QoS controllers)
5.6.5 Scalability in CPU use (end terminal and intermediate nodes)
5.6.6 Low latency in setup
5.6.7 Allow for low bandwidth consumption for signaling protocol
5.6.8 Ability to constrain load on devices
5.6.9 Highest possible network utilization

```
----------------------+-------------+-------------+-------------+
                      | host-to-net |   access    |   core      |
----------------------+-------------+-------------+-------------+
5.6.1                 |             |             |             |
```

5.6.2				
5.6.3				
5.6.4				
5.6.5				
5.6.6				
5.6.7				
5.6.8				
5.6.9				

5.7 Flexibility

5.7.1 Aggregation capability, including the capability to select and change the level of aggregation.
5.7.2 Flexibility in the placement of the QoS initiator
5.7.3 Flexibility in the initiation of re-negotiation (QoS change requests)
5.7.4 Uni / bi-directional reservation

	host-to-net	access	core
5.7.1			
5.7.2			
5.7.3			
5.7.4			

5.8 Security

5.8.1 The QoS protocol must provide strong authentication
5.8.2 The QoS protocol must provide means to authorize resource requests
5.8.3 The QoS signaling messages must provide integrity protection.
5.8.4 The QoS signaling messages must be replay protected.
5.8.5 The QoS signaling protocol must allow for hop-by-hop security.
5.8.6 The QoS protocol should allow identity confidentiality and location privacy.

5.8.7 The QoS protocol should prevent denial-of-service attacks against signaling entities.

5.8.8 The QoS protocol should support confidentiality of signaling messages.

5.8.9 The QoS protocol should provide hooks to interact with protocols that allow the negotiation of authentication and key management protocols.
5.8.10 The QoS protocol should provide means to interact with key management protocols.

	host-to-net	access	core
5.8.1			
5.8.2			
5.8.3			
5.8.4			
5.8.5			
5.8.6			
5.8.7			
5.8.8			
5.8.9			
5.8.10			

5.9 Mobility

5.9.1 Allow efficient QoS re-establishment after handover

	host-to-net	access	core
5.9.1			

5.10 Interworking with other protocols and techniques

5.10.1 Interworking with IP tunneling
5.10.2 The solution should not constrain either to IPv4 or IPv6

5.10.3 Independence from charging model
5.10.4 The QoS protocol should provide hooks for AAA protocols

----------------------+-------------+-------------+------------+

	host-to-net	access	core
5.10.1			
5.10.2			

5.10.3				
5.10.4				

5.11 Operational
5.11.1 Ability to assign transport quality to signaling messages

	host-to-net	access	core	
5.11.1				

7 References

[1] Kempf, J., "Dormant Mode Host Alerting ("IP Paging") Problem Statement", RFC 3132, June 2001.

[2] Chaskar, H., "Requirements of a QoS Solution for Mobile IP", draft-ietf-mobileip-qos-requirements-01.txt, Work in Progress, August 2001

[3] Manner. J., et al, "Mobility Related Terminology", draft-manner-seamoby-terms-02.txt, Work In Progress, July 2001.

[4] 3GPP, "General Packet Radio Service (GPRS); Service Description Stage 2 v 7.7.0", TS 03.60, June 2001

[5] 3GPP2, "Network Reference Model for cdma2000 Spread Spectrum System, revision B", S.R0005-B, May 2001

[6] Bradner, S., Mankin, A., "Report of the Next Steps in Signaling BOF", draft-bradner-nsis-bof-00.txt, Work in Progress, July 2001

[7] Partain, D., et al, "Resource Reservation Issues in Cellular Radio Access Networks", draft-westberg-rmd-cellular-issues-00.txt, Work in Progress, June 2001.

[8] YESSIR - YEt another Sender Session Internet Reservations, http://www.cs.columbia.edupingpan/projects/yessir.html

[9] Braden, R., Zhang, L., Berson, S., Herzog, A., Jamin, S., "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", IETF RFC 2205, 1997.

[10] Westberg, L., Jacobsson, M., Partain, D., Karagiannis, G.,

Oosthoek, S., Rexhepi, V., Szabo, R., Wallentin, P., "Resource Management in Diffserv Framework", Internet draft, work in progress, draft-westberg-rmd-framework-xx.txt, 2002.

[11] Kempf, J., McCann, P., Roberts, P., "IP Mobility and the CDMA Radio Access Network", IETF Draft, draft-kempf-cdma-appl-02.txt, Work in progress, September 2001.

8 Appendix: Scenarios/Use cases

In the following we describe scenarios, which are important to cover, and which allow us to discuss various requirements. Some regard this as use cases to be covered defining the use of a QoS signaling protocol.

8.1 Scenario: Terminal Mobility

The scenario we are looking at is the case where a mobile terminal (MT) changes from one access point to another access point. The access points are located in separate QoS domains. We assume Mobile IP to handle mobility on the network layer in this scenario and consider the various extensions (i.e., IETF proposals) to Mobile IP, in order to provide 'fast handover' for roaming Mobile Terminals. The goal to be achieved lies in providing, keeping, and adapting the requested QoS for the ongoing IP sessions in case of handover. Furthermore, the negotiation of QoS parameters with the new domain via the old connection might be needed, in order to support the different 'fast handover' proposals within the IETF.

The entities involved in this scenario include a mobile terminal, access points, an access network manager, communication partners of the MT (the other end(s) of the communication association).
From a technical point of view, terminal mobility means changing the access point of a mobile terminal (MT). However, technologies might change in various directions (access technology, QoS technology, administrative domain). If the access points are within one specific QoS technology (independent of access technology) we call this intra-QoS technology handoff. In the case of an inter-QoS technology handoff, one changes from e.g. a DiffServ to an IntServ domain, however still using the same access technology. Finally, if the access points are using different access technologies we call it inter-technology hand-off.

The following issues are of special importance in this scenario:

1) Handoff decision

- The QoS management requests handoff. The QoS management can decide to change the access point, since the traffic conditions of the new access point are better supporting the QoS requirements. The metric may be different (optimized towards a single or a group/class of users). Note that the MT or the network (see below) might trigger the handoff.

- The mobility management forces handoff. This can have several reasons. The operator optimizes his network, admission is no longer granted (e.g. emptied prepaid condition). Or another example is when the MT is reaching the focus of another base station. However, this might be detected via measurements of QoS on the physical layer and is therefore out of scope of QoS signaling in IP. Note again that the MT or the network (see below) might trigger the handoff.

- This scenario shows that local decisions might not be enough. The rest of the path to the other end of the communication needs to be considered as well. Hand-off decisions in a QoS domain, does not only depend on the local resource availability, e.g., the wireless part, but involves the rest of the path as well. Additionally, decomposition of an end-to-end reservation might be needed, in order to change only parts of it.

2) Trigger sources

- Mobile terminal: If the end-system QoS management identifies another (better-suited) access point, it will request the handoff from the terminal itself. This will be especially likely in the case that two different provider networks are involved. Another important example is when the current access point bearer disappears (e.g. removing the Ethernet cable). In this case, the QoS initiator is basically located on the mobile terminal.

- Network (access network manager): Sometimes, the handoff trigger will be issued from the network management to optimize the overall load situation. Most likely this will result in changing the base-station of a single providers network. Most likely the QoS initiator is located on a system within the network.

3) Integration with other protocols

- Interworking with other protocol must be considered in one or the other form. E.g., it might be worth combining QoS signaling between different QoS domains with mobility signaling at hand-over.

4) Handover rates

In mobile networks, the admission control process has to cope with far more admission requests than call setups alone would generate. For example, in the GSM (Global System for Mobile communications) case, mobility usually generates an average of one to two handovers per call. For third generation networks (such as UMTS), where it is necessary to keep radio links to several cells simultaneously (macro-diversity), the handover rate is significantly higher (see for example [11])

5) Fast reservations

Handover can also cause packet losses. This happens when the processing of an admission request causes a delayed handover to the new base station. In this situation, some packets might be discarded, and the overall speech quality might be degraded significantly. Moreover, a delay in handover may cause degradation for other users. In the worst case scenario, a delay in handover may cause the connection to be dropped if the handover occurred due to bad air link quality. Therefore, it is critical that QoS signalling in connection with handover be carried out very quickly.

6) Call blocking in case of overload

Furthermore, when the network is overloaded, it is preferable to keep reservations for previously established flows while blocking new requests. Therefore, the resource reservation requests in connection with handover should be given higher priority than new requests for resource reservation.

8.2 Scenario: Cellular Networks

In this scenario, the user is using the packet service of a 3rd generation cellular system, e.g. UMTS. The region between the End Host and the edge node connecting the cellular network to another QoS domain (e.g. the GGSN in UMTS or the PDSN in 3GPP2) is considered to be a single QoS domain [4][5].

The issues in such an environment regarding QoS include:

1) Cellular systems provide their own QoS technology with specialized parameters to co-ordinate the QoS provided by both the radio access and wired access network. For example, in a UMTS network, one aspect of GPRS is that it can be considered as a QoS technology; provisioning of QoS within GPRS is described mainly in terms of calling UMTS bearer classes. This QoS technology needs to be invoked with suitable parameters when a request for QoS is triggered by higher layers, and this therefore involves mapping the requested IP QoS onto these UMTS bearer classes. This request for resources might be triggered by IP signaling messages that pass across the cellular system, and possibly other QoS domains, to negotiate for network resources. Typically, cellular system specific messages invoke the underlying cellular system QoS technology in parallel with the IP QoS negotiation, to allocate the resources within the cellular system.

2) The placement of QoS initiators and QoS controllers (terminology in the framework given here). The QoS initiator could be located at the End Host (triggered by applications), the GGSN/PDSN, or at a node not directly on the data path, such as a bandwidth broker. In the second case, the GGSN/PDSN could either be acting as a proxy on behalf of an End Host with little capabilities, and/or managing aggregate resources within its QoS domain (the UMTS core network). The IP signaling messages are interpreted by the QoS controllers, which may be located at the GGSN/PDSN, and in any QoS sub-domains within the cellular system.

3) Initiation of IP-level QoS negotiation. IP-level QoS re-negotiation may be initiated by either the End Host, or by the network, based on current network loads, which might change depending on the location of the end host.

4) The networks are designed and mainly used for speech communication (at least so far).

Note that in comparison to the former scenario, the emphasis is much less on the mobility aspects, because mobility is mainly handled on the lower layer.

8.3 Scenario: UMTS access

The UMTS access scenario is shown in figure 3. The Proxy-Call State Control Function/Policy Control Function (P-CSCF/PCF) is the outbound SIP proxy of the visited domain, i.e. the domain where the mobile user wants to set-up a call. The Gateway GPRS Support Node (GGSN) is the egress router of the UMTS domain and connects the UMTS access network to the Edge Router (ER) of the core IP network. The P-CSCF/PCF communicates with the GGSN via the COPS protocol [4]. The User Equipment (UE) consists of a Mobile Terminal (MT) and Terminal Equipment (TE), e.g. a laptop.

```
                  +--------+
       +----------| P-CSCF |-------> SIP signaling
      /           +--------+
     / SIP            :
    :             +--------+   NSIS  +----------------+
    :             |  PCF   |---------| QoS Controller |
    :             +--------+         +----------------+
    :                 :
    :                 : COPS
    :                 :
  +----+          +--------+      +----+
  | UE |----------|  GGSN  |------| ER |
  +----+          +--------+      +----+
```

Figure 1: UMTS access scenario

In this scenario the GGSN has the role of Access Gate. According to 3GPP standardization, the PCF is responsible for the policy-based control of the end-user service in the UMTS access network (i.e. from UE to GGSN). In the current UMTS release R.5, the PCF is part of the P-CSCF, but in UMTS R.6 the interface between P-CSCF and PCF may evolve to an open standardized interface. In any case the PCF has all required QoS information for per-flow admission control in the UMTS access network (which it gets from the P-CSCF and/or GGSN). Thus the PCF would be the appropriate entity to host the functionality of QI, initiating the "NSIS" QoS signaling towards the core IP network. The PCF/P-CSCF has to do the mapping from codec type (derived from SIP/SDP signaling) to IP traffic descriptor. SDP extensions to explicitly signal QoS information [7] are useful to avoid the need to store codec information in the PCF and to allow for more flexibility and accurate description of the QoS traffic parameters. The PCF also controls the GGSN to open and close the gates and to configure per-flow policers, i.e. to authorize or forbid user traffic.

The QC is (of course) not part of the standard UMTS architecture. However, to achieve end-to-end QoS a QC is needed such that the PCF can request a QoS connection to the IP network. As in the previous example, the QC could manage a set of pre-provisioned resources in the IP network, i.e. bandwidth pipes, and the QC performs per-flow admission control into these pipes. In this way, a connection can be made between two UMTS access networks, and hence, end-to-end QoS can be achieved. In this case the QI and QC are clearly two separate entities.
This use case clearly illustrates the need for an "NSIS" QoS signaling protocol between QI and QC. An important application of such a protocol may be its use in the inter-connection of UMTS networks over an IP backbone.

8.4 Wired part of wireless network

A wireless network, seen from a QoS domain perspective, usually consists of three parts: a wireless interface part (the "radio interface"), a wired part of the wireless network (i.e., Radio Access Network) and the backbone of the wireless network, as shown in Figure 2. Note that this figure should not be seen as an architectural overview of wireless networks but rather as showing the conceptual QoS domains in a wireless network.

In this scenario, a mobile host can roam and perform a handover procedure between base stations/access routers. In this scenario the NSIS QoS protocol can be applied between a base station and the gateway (GW). In this case a GW can also be considered as a local handover anchor point. Furthermore, in this scenario the NSIS QoS protocol can also be applied either between two GWs, or between two edge routers (ER).

```
                          |--|
                          |GW|
   |--|                   |--|
   |MH|---                  .
   |--|  / |-------|        .
        /--|base   | |--|   .
           |station|-|ER|....
           |-------| |--|   .  |--| back- |--|  |---|
|----|

...|ER|.......|ER|..|BGW|.."Internet"..|host|
        -- |-------| |--|   .  |--| bone  |--|  |---|
|----|
   |--| \  |base   |-|ER|...     .
```

```
    |MH|  \ |station| |--|        .
    |--|--- |-------|             .          MH  = mobile host
                               |--|          ER  = edge router
      <---->                   |GW|          GW  = gateway
     Wireless link             |--|          BGW = border gateway
```

```
                                  ... = interior nodes
      <-------------------->
  Wired part of wireless network

 <------------------------------------------>
            Wireless Network
```

Figure 2. QoS architecture of wired part of wireless network

Each of these parts of the wireless network impose different issues to be solved on the QoS signaling solution being used:

* Wireless interface: The solution for the air interface link has to ensure flexibility and spectrum efficient transmission of IP packets. However, this link layer QoS can be solved in the same way as any other last hop problem by allowing a host to request the proper QoS profile.

* Wired part of the wireless network: This is the part of the network that is closest to the base stations/access routers. It is an IP network although some parts logically perform tunneling of the end user data. In cellular networks, the wired part of the wireless network is denoted as a radio access network.

  This part of the wireless network has different characteristics when compared to traditional IP networks:

  1. The network supports a high proportion of real-time traffic. The majority of the traffic transported in the wired part of the wireless network is speech, which is very sensitive to delays and delay variation (jitter).

  2. The network must support mobility. Many wireless networks are able to provide a combination of soft and hard handover procedures. When handover occurs, reservations need to be established on new paths. The establishment time has to be as short as possible since long establishment times for reservations degrade the performance of the wireless network. Moreover, for maximal utilization of the radio spectrum, frequent handover operations are required.

  3. These links are typically rather bandwidth-limited.

  4. The wired transmission in such a network contains a relatively high volume of expensive leased lines. Overprovisioning might therefore be prohibitively expensive.

5. The radio base stations are spread over a wide geographical area and are in general situated a large distance from the backbone.

* Backbone of the wireless network: the requirements imposed by this network are similar to the requirements imposed by other types of backbone networks.

Due to these very different characteristics and requirements, often contradictory, different QoS signalling solutions might be needed in each of the three network parts.

8.5 Scenario: Session Mobility

In this scenario, a session is moved from one end-system to another. Ongoing sessions are kept and QoS parameters need to be adapted, since it is very likely that the new device provides different capabilities. Note that it is open which entity initiates the move, which implies that the QoS initiator might be triggered by different entities.

User mobility (i.e., a user changing the device and therefore moving the sessions to the new device) is considered to be a special case within the session mobility scenario.

Note that this scenario is different from terminal mobility. Not the terminal (end-system) has moved to a different access point. Both terminals are still connected to an IP network at their original points.

The issues include:

1) Keeping the QoS guarantees negotiated implies that the end-point(s) of communication are changed without changing the reservations.

2) The trigger of the session move might be the user or any other party involved in the session.

8.6 Scenario: QoS reservations/negotiation from access to core network

The scenario includes the signaling between access networks and core networks in order to setup and change reservations together with potential negotiation.

The issues to be solved in this scenario are different from previous ones.

1) The entity of reservation is most likely an aggregate.

2) The time scales of reservations might be different (long living reservations of aggregates, rarer re-negotiation).

3) The specification of the traffic (amount of traffic), a particular QoS is guaranteed for, needs to be changed. E.g., in case additional flows are added to the aggregate, the traffic

specification of the flow needs to be added if it is not already included in the aggregates specification.

4) The flow specification is more complex including network addresses and sets of different address for the source as well as for the destination of the flow.

8.7 Scenario: QoS reservation/negotiation over administrative boundaries

Signaling between two or more core networks to provide QoS is handled in this scenario. This might also include access to core signaling over administrative boundaries. Compared to the previous one it adds the case, where the two networks are not in the same administrative domain. Basically, it is the inter-domain/inter provider signaling which is handled in here.

The domain boundary is the critical issue to be resolved. Which as various flavors of issues a QoS signaling protocol has to be concerned with.

1) Competing administrations: Normally, only basic information should be exchanged, if the signaling is between competing administrations. Specifically information about core network internals (e.g., topology, technology, etc.) should not be exchanged. Some information exchange about the "access points" of the core networks (which is topology information as well) may need to be exchanged, because it is needed for proper signaling.

2) Additionally, as in scenario 4, signaling most likely is based on aggregates, with all the issues raise there.

3) Authorization: It is critical that the QoS initiator is authorized to perform a QoS path setup.

4) Accountability: It is important to notice that signaling might be used as an entity to charge money for, therefore the interoperation with accounting needs to be available.

8.8 Scenario: QoS signaling between PSTN gateways and backbone routers

A PSTN gateway (i.e., host) requires information from the network regarding its ability to transport voice traffic across the network. The voice quality will suffer due to packet loss, latency and jitter. Signaling is used to identify and admit a flow for which these impairments are minimized. In addition, the disposition of the signaling request is used to allow the PSTN GW to make a call routing decision before the call is actually accepted and delivered to the final destination.

PSTN gateways may handle thousands of calls simultaneously and there may be hundreds of PSTN gateways in a single provider network. These numbers are likely to increase as the size of the network increases. The point being that scalability is a major issue.

There are several ways that a PSTN gateway can acquire assurances that a network can carry its traffic across the network. These include:

1. Over-provisioning a high availability network.
2. Handling admission control through some policy server that has a global view of the network and its resources.
3. Per PSTN GW pair admission control.
4. Per call admission control (where a call is defined as the 5 tuple used to carry a single RTP flow).

Item 1 requires no signaling at all and is therefore outside the scope of this working group.

Item 2 is really a better informed version of 1, but it is also outside the scope of this working group as it relies on a particular telephony signaling protocol rather than a packet admission control protocol.

Item 3 is initially attractive as it appears to have reasonable scaling properties, however, its scaling properties only are effective in cases where there are relatively few PSTN GWs. In the more general case were a PSTN GW reduces to a single IP phone sitting behind some access network, the opportunities for aggregation are reduced and the problem reduces to item 4.

Item 4 is the most general case. However, it has the most difficult scaling problems. The objective here is to place the requirements on Item 4 such that a scalable per-flow admission control protocol or protocol suite may be developed.

The case where per-flow signaling extends to individual IP end-points allows the inclusion of IP phones on cable, DSL, wireless or other access networks in this scenario.

Call Scenario

A PSTN GW signals end-to-end for some 5 tuple defined flow a bandwidth and QoS requirement. Note that the 5 tuple might include masking/wildcarding. The access network admits this flow according to its local policy and the specific details of the access technology.

At the edge router (i.e., border node), the flow is admitted, again with an optional authentication process, possibly involving an external policy server. Note that the relationship between the PSTN GW and the policy server and the routers and the policy server is outside the scope of NSIS. The edge router then admits the flow into the core of the network, possibly using some aggregation technique.

At the interior nodes, the NSIS host-to-host signaling should either be ignored or invisible as the Edge router performed the admission control decision to some aggregate.

At the inter-provider router (i.e., border node), again the NSIS host-to-host signaling should either be ignored or invisible as the Edge router has performed an admission control decision about an aggregate across a carrier network.

8.9 PSTN trunking gateway

One of the use cases for the NSIS signaling protocol is the scenario of interconnecting PSTN gateways with an IP network that supports QoS.
Four different scenarios are considered here.

1. In-band QoS signaling is used. In this case the Media Gateway (MG) will be acting as the QoS Initiator and the Edge Router (ER) will be the QoS Controller. Hence, the ER should do admission control (into pre-provisioned traffic trunks) for the individual traffic flows. This scenario is not further considered here.
2. Out-of-band signaling in a single domain, the QoS Controller is integrated in the MGC. In this case no NSIS protocol is required.
3. Out-of-band signaling in a single domain, the QoS Controller is a separate box. In this case NSIS signaling is used between the MGC and the QoS Controller.
4. Out-of-band signaling between multiple domains, the QoS Controller (which may be integrated in the MGC) triggers the QoS Controller of the next domain.

When the out-of-band QoS signaling is used the Media Gateway Controller (MGC) will be acting as the QoS Initiator.

In the second scenario the voice provider manages a set of traffic trunks that are leased from a network provider. The MGC does the admission control in this case. Since the QoS Controller acts both as a QoS Initiator and a QoS Controller, no NSIS signaling is required. This scenario is shown in figure 1.

```
   +-------------+    ISUP/SIGTRAN     +-----+              +-----+
   | SS7 network |---------------------| MGC |--------------| SS7 |
   +-------------+             +-------+-----+---------+    +-----+
         :                    /           :             \
         :                   /            :              \
         :                  /    +--------:---------+     \
         :          MEGACO /    /         :           \    \
         :                /    /       +-----+         \    \
         :               /    /        | NMS |          \    \
         :              /     |        +-----+          |     \
         :              :     |                         |     :
  +--------------+  +----+    |   bandwidth pipe (SLS)  |  +----+
```

```
 | PSTN network |--| MG |--|ER|======================|ER|-| MG |--
 +--------------+  +----+     \                        /   +----+
                               \     QoS network     /
                                +--------------------+
```

Figure 1: PSTN trunking gateway scenario

In the third scenario, the voice provider does not lease traffic trunks in the network. Another entity may lease traffic trunks and may use a QoS Controller to do per-flow admission control. In this case the NSIS signaling is used between the MGC and the QoS Controller, which is a separate box here. Hence, the MGC acts only as a QoS Initiator. This scenario is depicted in figure 2.

```
      +-------------+    ISUP/SIGTRAN     +-----+              +-----+
      | SS7 network |---------------------| MGC |--------------| SS7 |
      +-------------+             +-------+-----+---------+    +-----+
            :                    /           :             \
            :                   /         +-----+           \
            :                  /          | QC  |            \
            :                 /           +-----+             \
            :                /               :                 \
            :               /       +--------:----------+       \
            :       MEGACO :       /         :           \       :
            :              :      /       +-----+         \      :
            :              :     /        | NMS |          \     :
            :              :     |        +-----+          |     :
            :              :     |                         |     :
     +--------------+  +----+    |   bandwidth pipe (SLS)  |  +----+
     | PSTN network |--| MG |--|ER|======================|ER|-| MG |--
     +--------------+  +----+     \                       /   +----+
                                   \     QoS network     /
                                    +-------------------+
```

Figure 2: PSTN trunking gateway scenario

In the fourth scenario multiple transport domains are involved. In the originating network either the MGC may have an overview on the resources of the overlay network or a separate QoS Controller will have the overview. Hence, depending on this either the MGC or the QoS Controller of the originating domain will contact the QoS Controller of the next domain. The MGC always acts as a QoS Initiator and may also be acting as a QoS Controller in the first domain.

8.10 Scenario: Application request end-to-end QoS path from the network

This is actually the most easy case, nevertheless might be most often used in terms of number of users. So multimedia application requests a guaranteed service from an IP network. We assume here that the application is somehow able to specify the network service. The characteristics here are that many hosts might do it, but that the requested service is low capacity (bounded by the access line). Additionally, we assume no mobility and standard devices.

9 Acknowledgments

Quite a number of people have been involved in the discussion of the draft, adding some ideas, requirements, etc. We list them without a guarantee on completeness: Changpeng Fan (Siemens), Krishna Paul (NEC), Maurizio Molina (NEC), Mirko Schramm (Siemens), Andreas Schrader (NEC), Hannes Hartenstein (NEC), Ralf Schmitz (NEC), Juergen Quittek (NEC), Morihisa Momona (NEC), Holger Karl (Technical University Berlin), Xiaoming Fu (Technical University Berlin), Hans-Peter Schwefel (Siemens), Mathias Rautenberg (Siemens), Christoph Niedermeier (Siemens), Andreas Kassler (University of Ulm), Ilya Freytsis.

Some text and/or ideas for text, requirements, scenarios have been taken from a draft written by the following authors: David Partain (Ericsson), Anders Bergsten (Telia Research), Marc Greis (Nokia), Georgios Karagiannis (Ericsson), Jukka Manner (University of Helsinki), Ping Pan (Juniper), Vlora Rexhepi (Ericsson), Lars Westberg (Ericsson), Haihong Zheng (Nokia). Some of those have actively contributed new text to the draft as well.

Another draft impacting this draft has been written by Sven Van den Bosch, Maarten Buchli, and Danny Goderis. These people contributed also with new text.

10 Author's Addresses

Marcus Brunner (Editor)
NEC Europe Ltd.
Network Laboratories
Adenauerplatz 6
D-69115 Heidelberg
Germany
E-Mail: brunner@ccrle.nec.de (contact)

Robert Hancock, Eleanor Hepworth
Roke Manor Research Ltd
Romsey, Hants, SO51 0ZN
United Kingdom
E-Mail: [robert.hancock|eleanor.hepworth]@roke.co.uk

Cornelia Kappler
Siemens AG
Berlin 13623
Germany
E-Mail: cornelia.kappler@icn.siemens.de

Hannes Tschofenig
Siemens AG

Otto-Hahn-Ring 6
81739 Munchen
Germany
Email: Hannes.Tschofenig@mchp.siemens.de

Open Issues/To Dos

1) (OPEN) add Scenarios

Do we need to add, remove, or change the scenarios?

- added scenario on QoS signalling between PSTN gateways and backbone routers
- added: Application request end-to-end QoS path from the network

We can what ever scenario we want. The more the better to understand the issues. Nevertheless, we have to take care that we are future prove as well.

2) (OPEN) Sender/receiver initiation

What is the requirement concerning data sender or data receiver or both to initiate a QoS request.

Terminology text added

open issue, what is a potential req (currently we say "both must be possible")

Proposals:

1)should be optimized for sender initiated

2) remove the requirement, because it is not relevant if we allow

for third party QoS initiators
3) SHOULD support sender initiated, MAY support reciever initiated

Issues:

- does it matter who pays?

- for sender initiated, can we support implicit signaling (bundling the QoS requests with other signaling - registration, etc.)?

- For reciever initiated, do we need protection against spamming - how do we protect against unwanted changes?

3) (CLOSED) Draft organization

The proposed changes include
- put all the scenarios into an appendix
- In Section 6 add text describing 3 different "parts of the network"
     -Host to first hop
     -access network
     -core networks

  better names are welcome, but I don't want to be religious about it

- Prioritize the requirements according to the "parts of the network" (This means the the tables in Section 6 of the current draft will get three colums with the MUST, SHOULDs, and MAYs for each requirement

4) (OPEN) MUSTs, SHOULDs, MAY needs discussion

5) (CLOSED) Framework text.
The figures have been removed, because they seamed to be misleading. the text has been revisited. I regard the issue closed until we have a clear picture about what the NSIS framework draft is about.

6) (CLOSED) The requirement organization
I heard some voices on the list that the grouping should be more along the lines of host-to-edge, edge to edge etc.
So far I have not changed it, because I though that the requirements heavily depend on the scenario we are looking at.

closed, by the change in the draft organisation (issue 3)

7) (OPEN) Hemant Chaskar: Section 3.1, items 1) Handoff decision and 2) Trigger sources: The handoff decision and trigger sources should be out of scope of NSIS. NSIS should rather focus only on "establishing" QoS along the packet path after handoff.

needs more WG discussion, potentially even cross-WG

8) (OPEN) bi-directional data path setup with one QoS request
I have not seen consensus on whether to require bi-directional data path setup with QoS.

Q: How can we actually perform bi-directional reservations when the forward and reverse paths are not reciprocal, with respect to routing topology and routing policy of network domains between sender and receiver?

A: bi-directional data path setup does not need to use the same return path as the forwarding path. The only requirement to achieve a bi-directional reservation is that the sender for the forwarding path is also the receiver for the return path and that the receiver for the forwarding path is also the sender for the return path.

- The need to ensure that the return path is the same as the forwarding path is one of the problems with RSVP, particularly in a mobile environment.

9) (CLOSED) Potential requirement: must be implementable in user space (on end hosts)

has not been included in the req list because it seams to be implementation specific.

10) (CLOSED) Potential requirement: must provide support for globally defined services as well as private services (Ruediger)

replaced by issue 17 and 18, closed

11) (CLOSED) Potential requirement: Flexibility in the granularity of reservation (I don't remember who brought it up, but I assume it refers to the flexibility in terms of what size the flow has. Where size can be bandwidth etc.)

The assumption that QoS classes as well as service definitions are out of scope for this draft, also the flexibility is.

12) (CLOSED) text replacing scalability reqs

"The nsis architecture should give the ability to constrain the load (CPU load, memory space, signaling bandwidth consumption and signaling intensity) on devices where it is needed. This can be achieved by many different methods, for example message aggregation, by ignoring signaling message, header compression or minimizing functionality. The architecture may choose any of these methods as long as the requirement is met."

Editor: added the draft text, but did not remove scalability reqs

13) (CLOSED) add operator req "Ability to assign transport quality to signaling messages"
"The nsis architecture should allow the network operator to assign

the nsis protocol messages a certain transport quality. As signaling opens up for possible denial-of-service attacks, this requirement gives the network operator a mean, but also the obligation, to trade-off between signaling latency and the impact (from the signaling messages) on devices within his/her network. From protocol design this requirement states that the protocol messages should be detectable, at least where the control and assignment of the messages priority is done."

text has been added

14) (OPEN, dependend on resolution of bi-directional) proposal to add "support grouping of microflows (possibly only for feedback)" "As a consequence of the optimization for the interactive multimedia services, the signaling should allow one unique request for several micro flows having the same origination and destination IP addresses. This is usually the case for multimedia SIP calls where the voice and video micro flows follow the same path. This grouping of requests allows optimization of the QoS processing. Note that this may be detrimental for the call setup time. The use of grouping for microflows may be restricted to teardown and/or notification messages when call setup time is a concern."

open issue: first resolve the bi-directional issue which is somewhat related, because it seams to be an optimization as well

Should not be restrict to teardown and/or notification, it might be useful also for the procedure that refreshes reservation states

15) (CLOSED) Support for preemption of sessions
-might play into the fault/ error handling case
-is regarded as service-specific, whether existing sessions can be pre-empted
Conclusion: it is network policy to determine how to do pre-emption, not a protocol issue.

16) (OPEN) Req: 5.1.9 change provisioning into better term, since different people understand different thing with provisioning

open action for Anders

17) (CLOSED) add assumption that QoS classes/service definitions are already known to all the parties involved in signaling before hand (before a signalling session even starts

added text in Section 4.1

18) (CLOSED) add exclusion of methods, protocols, and ways to express QoS
Even so, this might be covered by saying that we are independent of QoS classes and service description etc. (see issue 17), I added two points to the exclusion Section 4.2.

Implications: issue 20, 23,

19) (CLOSED) remove req 5.2.5 IP fragmentation

20) (CLOSED) remove req 5.3.2 Ability to signal life-time of a reservation

is regarded service-specific therefore part of the service description

added some reservation life time text service description assumption text and removed the req

21) (CLOSED) remove req 5.5.4 Aggregation method specification

Concerns
-QI not able to know the impact of aggregation
-to far down the implementation/ service definition road
-leave it to the provider how a service is realized

removed

22) (CLOSED) remove 5.3.7 Automatic notification on available resources not been granted before

regarded to complex and is heavily dependend on the service description

removed

23) (CLOSED) remove 5.5.3 Simple mapping to lower-layer QoS provisioning parameters

this heavily depends on service definition and therefore is out of scope of this document

removed

24) (CLOSED) Replacing req 5.3.6 "Feedback about the actually received level of QoS guarantees" with two requirements: 1) the feedback of a request MUST include yes and no (MUST respond yes or no) 2) in case of no it MAY include an opaque service-specific information about what would be possible

It is still only one requirement, but the text has been replaced.

25) (CLOSED) remove req 5.10.3 Combination with Mobility management

However the integration should not be a priori excluded, there is explicitly no statemant about independence of mobility management.

There is more discussion for the mobility case needed anyway.

26) (OPEN) interaction of NSIS with seamoby (context transfer and

CAR discovery)

27) (CLOSED) remove req 5.5.10 QoS conformance specification

Motivation: this heavily depends on the service definition and is therefore out of scope

removed

28) (OPEN) new requirement on "asynchronous events from the network"

The content of the message might be very service specific, but the protocol support for asynchronous events from the network might be a valuable requirement. We have something about notification in case of errors/failures.

29) (OPEN) NSIS in case of handovers
The whole mobility area needs to be defined

30) (CLOSED) remove 5.1.7 Avoid modularity with large overhead (in various dimensions)

removed because it seams to be obvious

31) (CLOSED) remove 5.1.8 Possibility to use the signaling protocol for existing local technologies

It is contradictory to 5.1.9 and the intention behind the requirement is covered by the requierement that the QoS controller can be placed wherever needed.

32) (CLOSED) add assumption: there are means for discovery of nsis entities in order to know the signaling peers (solutions include static configuration, or automatically discovered etc.)

33) (CLOSED) add req " highest possible network utilization"
"There are networking environments that require high network utilization for various reasons, and the signaling protocol should to its best ability support high resource utilization while maintaining appropriate QoS.

In networks where resources are very expensive (as is the case for many wireless networks), efficient network utilization is of critical financial importance. On the other hand there are other parts of the network where high utilization is not required.
"

req added

34) (CLOSED)_difference between "UMTS access scenario" "cellular network scenario", and "Wired part of wireless network" (Section 8.2, 8.3, and 8.4)

all three are included.

The only common point between the three scenarios is that they are related to cellular networks. Section 8.4 is introducing the scenario used in the radio access network of cellular networks.

Sections 8.2 and Section 8.3 are discussing other parts of the cellular network.

35) (CLOSED) difference between the two PSTN gateway scenarios (Section 8.8 and 8.9)

currently both are included, they might be merged, sionce one seams to be more general than the other

36) (OPEN) req "Independence of reservation identifier"
issue here is that this might only be valuable in mobile environments, and complicate the protocol for other environemnts.

there are related issues (37,38,

37) (OPEN) ownership of a reservation

The issue here is that a known party owns reservations done in the network. (which might include that the party also pays). The question arose who is allowed to tear-down, receive asynchronous notifications in case of network initiated tear-down, etc.

This also relates to how certain service granted is named/identified.

38) (OPEN) definition of security threats

39) (OPEN) simplify security requirements section

40) (OPEN) add mobility related requirements

41) (CLOSED) remove req 5.5.1 Mutability information on parameters removed because it is service-specific

42) (OPEN) add an assumption that QoS nmonitoring is application-specific and with it out of scope of the WG

43) (OPEN) asynchronous notification of QoS Initiator, Controller, Receiver, there are security issues related. Basically, an ownership issue. Nevertheless, an asynch notifcation in case of an error, network failure etc. is specifically in areas, where longer lived sessions are setup, essential in order to notify upper layes (appluications etc. as well.

44) (OPEN) req 5.1.2 resource availability info on request come back to it as soon as we have a more clear idea about service description issue

45) (OPEN) 5.3.4 Possibility for automatic re-setup of resources after recovery

- more thoughts in failure conditions potentially
- better text
- operation under overload

plays into issue 46)

46) (OPEN) we need multiple scenario for failure and recovery cases to derive requirements. Or a list failre cases might be a start as well.

47) (OPEN) traffic engineering and route pinning
I assume this would result in operational type of requirements
Opinions on that?

48) (CLOSED) req 5.5.5 remove Multiple levels of detail

"The QSC should allow for multiple levels of detail in description. (Motivation: someone interpreting the request can tune its own level of complexity by going down to more or less levels of detail. A lightweight implementation within the core could consider only the coarsest level.)"

removed, because it is service-specific

49) (CLOSED) remove req 5.5.9 Signaling must support quantitative, qualitative, and relative QoS specifications

removed because it is service-specific

50) (CLOSED) req 5.5.6 remove Ranges in specification

The QSC should allow for specification of minimum required QoS and/or desirable QoS. (Motivation: The QoS Service Classes should allow for ranges to be indicated, to minimize negotiation latency and suppress error notifications during handover events.)

removed, is service specific

51) (CLOSED) remove 5.1.6 Avoid duplication of [sub]domain signaling functions

we might use the requirement text somewhere else:

Heading: Avoid duplication of [sub]domain signaling functions

The specification of the NSIS signaling protocol should be optimized to avoid duplication of existing [sub]domain QoS signaling and to minimize the overall complexity. (Motivation: we don't want to introduce duplicate feedback or negotiation mechanisms, or complicate the work by including all possible existing QoS signaling in some form. The function will be placed in the new part if it has to be end-to-end, universal to all network types ('simple/lightweight'), or if it has to be protected by upper layer security mechanisms.)

The point here is that the QoS technology (lower layer stuff) gets re-used unchanged, and we have new signaling above it. But, in many cases the local QoS technology will contain equivalent functions to the NSIS-required ones, just in a technology specific form. Examples of these functions would be error/QoS violation notifications, ability to query for resources and so on. So, there is a danger that our 'lightweight' signaling ends up trying to carry all this information all over again, and (even worse) that the initiator/controller functions have to weigh up nearly equivalent information coming from two directions. However, the basic problem here is that the boundary between new and re-used stuff is pretty shaky. The requirement is trying to scope our problem (a) to eliminate the potential overlap, and (b) to keep the new NSIS stuff simple.

However, we are aware that it is very difficult to judge what is duplicated, if we want to run the protocol in various environments.

52) (OPEN) New requirement: interaction with policy
this most likely is covered by an opaque token for authentication dependency on security changes

53) (OPEN) Section 5.3. Error handling

Comments:
1) notification of user in case of unrecoverable errors (has been done by notification requirement, or will be done by asynch notification, issue 43)
A description of both types of errors (recoverable, unrecoverable) are listed in Section 5.3.4.

2) hop-by-hop? OR right to the end?

3) What is potential value to notify about recoverable errors?
Proposal: not hop by hop, but QoS controller to QoS initiator

54) (CLOSED) add req 5.1.17. to assumption "Identification requirement"
assumption say that the discovery of QI, QC, QR is out-of-scope of the draft

55) (CLOSED) add from draft-partain-nsis-requirements-00.txt req 5.2.2. Allow local QoS information exchange between two border nodes

"The QoS signalling protocol must be able to exchange local QoS information between edge nodes. Local QoS information might, for example, be IP addresses, severe congestion notification, notification of succesful or erroneous processing of QoS signalling messages at one border node.

In some domains, the NSIS QoS signalling protocol MAY carry identification of the ingress and egress edge between the ingress-egress edges. However, the identification of edges should not be visible to the end host and only applies within one QoS administrative domain.

"

Comments:
- service mapping is more service-specific (layering,tunneling)
- the scenario to look at is a complicated service description -> in part of the network you want to change the message to something more easy, and at the other end go back to the more complicated part.
-QI being everywhere might be enough
-and we have already a requirement saying that intermediate node MUST be able to add/remove domain-specific information to/from signaling messages

56) (CLOSED) add req 5.3.1.3 of draft-partain-..-00
-already added a req to the scalability section (issue ???), which has been provided by Anders

57) (CLOSED) potentially better title for text from issue 56) e.g. (ôminimal impact on coreö)

58) (CLOSED) add req 5.3.2 from draft-partain-...-00

- the fast establishment req is handled by the low setup latency req, and the scalability in handover req

- added the text to the teminal mobility scenario

- added text " time scale (e.g., handover in mobile environments)," to req

59) (OPEN) add req: ability to deal with severe congestion (req 5.3.4 of draft-partain-..-00

issues are:
- occurs in a highly utilised network and if it is not solved very fast then the network performance will quickly collapse
- deos it belong to failure recovery (I would assume from a service point of view this is failure
- hop by hop problem (issue from Jorge)
- What difference does it make (from the QoS perspective) if the provided QoS degraded due to hardware failure on a device or due to congestion caused by failures on some other devices? What is required from the protocol is to signal this failure to other participants (QCs or QI) in the hope that they can do something meaningful (e.g. re-routing) to correct the problem or tear down the flow.

60) (CLOSED) add req 5.4.3. from draft-partain-...-00 "Allow efficient QoS re-establishment after handover"

"Handover is an essential function in wireless networks. After handover, QoS may need to be completely or partially re-established due to route changes. The re-establishment may be requested by the mobile node itself or triggered by the access point that the mobile node is attached to. In the first case, the QoS signalling should allow efficient QoS re-establishment after handover. Re-establishment of QoS after handover should be as quick as possible so that the mobile node does not experience service interruption or QoS degradation. The re-establishment should be localized, and not require end-to-end signalling, if possible."

- most likely it is already cover, please check again, whether there is something missing
- added it again under the mobility requiremments

61) (OPEN) add req: 6.1.8 from draft-bucheli-...-00 on multicast "Multicast consideration should not impact the protocol complexity for unicast flows. Multicast support is not considered as a priority, because the targeted interactive multimedia services are mainly unicast. For this reason, if considered in the solution, multicast should not bring complexity in the unicast scenario."

Opinions?

----
starting from -02 version
----

62) (OPEN) Request to add VPN scenario
- Related to issue 1)
- Difference of VPN scenario compared to what we already have is missing

63) (CLOSED) added Sven Van den Bosch, Maarten Buchli, and Danny Goderis to acknowledgement section.

64) (OPEN) Request to add req: Backwards compatibility
A later version of an NSIS protocol must be backwards compatible with earlier versions of an NSIS protocol.

65) (OPEN) Request to add req: Unexpected situations and error restistance
An NSIS protocol must define behaviour of NSIS signaling units during unexpected situations. Unexpected situtions are unknown messages, parameters and parameter settings as well as receiption of unexpected messages (e.g. a "Reservation Confirmation" without prior "Reservation Request").

Related to Open issues (53) and requirement 5.3.4.
This requirement is emphasizing to many details that might not be necessary

Req 5.3.4 refers to behaviour in the case of problems in the data plane. My suggestion here is about unexpected events/errors in the

control plane. If you think that this point carries to many details,
let's split it up in several individual requirements.

66) (OPEN) Request to add req: Default behaviour
An NSIS protocol must define default behaviours and parameter settings wherever applicable.

67) (OPEN) Request to add req: Extendability
An NSIS protocol must provide means to enhance a protocol with future procedures, messages, parameters and parameter settings.

This was refering mostly to the service specific part of the protocol.
could be a part of the modularity requirement 5.1.3

68) (OPEN) Request to add req: Preventation of stale state
An NSIS signalling protocol must provide means for an NSIS signaling unit to discover and remove local stale state. This may for example be done by means like soft state and periodic flooding or by a polling mechanism and hard state signaling.

Might already be covered in other requirements, could also be that the solutions known are solutions for different problems. I think distributed garbage collection could also be a solution.

69) (OPEN) Request to add req: Reliable Communication
NSIS signaling procedures, connectivity between units involved in NSIS signaling as well as the basic transport protocol used by NSIS must provide a maximum of communication reliability. Procedures must define how an NSIS signaling systems behaves if some kind of request it sent stays without answer (this could require e.g. be timers, number of message retransmits and release messages).
An NSIS signaling unit must be able to check its connectivity to an adjacent NSIS signaling unit at any time (this requirement must however not result in a DoS attack tool - the frequency of these checks must be limited, and flow control may be useful).
The basic transport protocol to be used between adjacent NSIS units must ensure message integrity and reliable transport.

MUST/SHOULD ensure error- and loss free transmission of signaling information.

Do we really require this? Isn't this a soft state versus hard state issue?

70) (OPEN) Request to add req: Smooth breakdown
A unit participating in NSIS signaling must no cause further damage to other systems involved in NSIS signaling when it has to go out of service.

71) (CLOSED) Changed text "5.6.8: Ability to constrain load on devices" to

The NSIS architecture should give the ability to constrain the load (CPU load, memory space, signaling bandwidth consumption and signaling intensity) on devices where it is needed. This can be achieved by many different methods. Examples, and this are only examples, include message aggregation, by ignoring signaling message, header compression, or minimizing functionality. The framework may choose any method as long as the requirement is met.

72) (OPEN) request to add "Error notification and error location"

"An NSIS signaling node rejecting or releasing a reservation must indicate its identity. NSIS signalling should indicate why a requested resource is not or no longer available. "

Compared to 5.3.4 this is about problems on the control plane

----

Change Log Version 01 -> 02

- added issues 62-72

- added some discussion text to open issues

- req " highest possible network utilization" added (issue 33, closed)

- issues closed: 34 (UMTS scenarios), 35 (PSTN gatway scenarios),

- removed req "Avoid duplication of [sub]domain signaling functions", issue 51

- Section 5.3.4: added explanation of recoverable and unrecoverable errors (issue 53)

- added the following requirement: (closed issue 55) Allow local QoS information exchange between nodes of the saeme administrative domain

The QoS signaling protocol must be able to exchange local QoS information between QoS controllers located within one single domain. Local QoS information might, for example, be IP addresses, severe congestion notification, notification of successful or erroneous processing of QoS signaling messages.
In some cases, the NSIS QoS signalling protocol may carry identification of the QoS controllers located at the boundaries of a domain. However, the identification of edge should not be visible to the end host (QoS initiator) and only applies within one QoS administrative domain.

- closed issue 57: add text about "Minimal impact on interior (core) nodes" to requirement 5.6.8 "Ability to constrain load on devices"

- added requirement "Allow efficient QoS re-establishment after handover", closed issue 60.

- changed text in 5.3.2