Internet Engineering Task Force
INTERNET-DRAFT
Expires July 2004

A. Bader
L. Westberg
Ericsson

G. Karagiannis
University of Twente

February 2004

RMD (Resource Management in Diffserv) QoS-NSLP model
draft-bader-rmd-qos-model-00.txt

Status of this memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts. Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet- Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (2002). All Rights Reserved.

Abstract

This draft describes a local QoS model, denoted as Resource Management in Diffserv (RMD) QoS model, for NSIS that extends the IETF Differentiated Services (Diffserv) architecture with a scalable admission control and resource reservation concept. The specification of this QoS model includes a description of its QoS parameter information, as well as how that information should be treated or interpreted in the network.

Table of Contents

1. Introduction

The Quality of Service NSIS Signaling Layer Protocol (QoS-NSLP) establishes and maintains states at nodes along the path of a data flow for the purpose of providing or querying forwarding resources for that flow [QoS NSLP]. The QoS NSLP separates the actual description of resources from the QoS signalling protocol used to transport them. It uses interchangeable QoS Models that allow the resource specification to be performed in various ways, and to provide different processing models (including reserve/commit models, measurement based models, etc).

A QoS model is a defined mechanism for achieving QoS as a whole. The specification of a QoS model includes a description of its QoS parameter information, as well as how that information should be treated or interpreted in the network. In that sense, the QoS model goes beyond the QoS-NSLP protocol level in that it could also describe underlying assumptions, conditions and/or specific provisioning mechanisms appropriate for it.

The actual resources that are required and are specific to the QoS model are specified in QSpec objects. Besides resource description, the QSpec objects may also contain QoS Model specific control information. In NSIS there is no restriction on the number of supported QoS models. QoS models may be local (private to one network), implementation/vendor specific, or global (implementable by different networks and vendors).

This draft describes a scalable edge-to-edge local QoS model, denoted as Resource Management in Diffserv (RMD) QoS model. This QoS model extends the IETF Differentiated Services (Diffserv) architecture with a scalable admission control and resource reservation concept.

Developing RMD QoS model is motivated by the need of a lightweight NSIS design that provides QoS for large number of flows and in the same time ensures fast message forwarding, for example in a core network or in a mobile radio access network (see Section 10 of [Brun03]). On the other hand these networks may provide services that allow the use of simplified resource reservation and transport schemes. The aim is to define an NSIS QoS model with simplified functionality, which may be implemented by dedicated network processors.

For example in an MPLS (Multi Protocol Label Switching) core network labels switched paths provide an aggregated bandwidth guarantee, as

well as transport and routing between edge routers. A simplified NSIS operation together with MPLS can provide end-to-end or edge-to-edge QoS without the need of significant over-dimensioning the Label Switched Paths (LSP-s).

A mobile radio access network can be characterized as a network that supports frequently changing reservation states (due to mobility), scarced bandwidth (due to the radio links) and simple network topologies (tree, star, ring and combination of these). 3G networks use bearer classes, i.e., Radio Access Bearers, to deliver user traffic, which are characterized by simple traffic descriptors. QoS in the IP transport network has to be provided for such RAB connections and the admission control is based on the number of connections per RAB type. Due to the frequently modification of aggregated reservations, lightweight signaling and simple reservation schemes based on e.g. bandwidth units, has to be applied. Moreover, fast release of unused resources (explicit release)is required. Furthermore, a fast fault handling algorithm that in case of link or node failure, re-routes traffic in a short time, from one route to an alternative one, without terminating the ongoing flows is also an essential requirement.

The local QoS model described here is based a on a simple, hop-by-hop probing mechanism. When a new flow arrives with some requested resources (typically bandwidth), each router checks the available resources on the future path of the flow. This is basically done by sending a probe packet through a path which indicates the required resources. If an intermediate router cannot accommodate the new request, then this situation is indicated by marking a single bit in the packet.

This QoS model is based the original concept of Resource Management in Diffserv (RMD) framework [RMD]. In RMD, scalability is achieved by separating a complex reservation mechanism used in the edge nodes of a domain from a much simpler reservation mechanism needed in the interior nodes of this domain. In particular, it is assumed that edge nodes of a Diffserv domain support per-flow QoS-NSLP states in order to provide QoS guarantees for each flow. Interior nodes between edge nodes use only one aggregated reservation NSLP state per traffic class or no states at all. In this QoS model interior routers do not store NTLP states, therefore, the NSLP messages used by these routers are transported by UDP/IP or IP only (i.e., NTLP datagram mode). This solution allows fast processing of signalling messages and makes it possible to handle large number of flows in the interior nodes.

Two basic operation modes are described: a measurement based admission control and a reservation based admission control. The measurement-based algorithm uses the requested and available resources as input to query the aggregated reservation state per traffic class in the interior nodes. The advantage of measurement based resource management protocols is that they do not require explicit reservation or release. Moreover, when the user traffic is variable, measurement based admission control could provide higher network utilization than, e.g., peak-rate reservation.

In case of the reservation-based method, each node in the domain maintains one reservation state per traffic class. The reservation is done in resource units. These resources are requested dynamically per PHB (Per Hop Behavior) and reserved on demand in all nodes in the communication path from an ingress node to an egress node.

1.1. Definitions/Terminology

The terminology defined in [QoS-NSLP] and [RFC2475] applies to this draft. In addition, the following terms are used:

- QNE - an NSIS Entity (NE), which supports the QoS-NSLP.
- QoS-NSLP stateful operation - mode of operation where per-flow reservation states are created, maintained and used.
- QoS-NSLP reduced-state operation - mode of operation where reservation states with a coarser granularity (e.g. per-class) are created, maintained and used.
- QoS-NSLP stateless operation - mode of operation where reservation state is not needed and not created.
- reduced state QNE - a QNE that supports the QoS NSLP reduced state operation.
- RMD domain - Administrative domain where an QoS-NSLP protocol signals for a resource or set of resources that are using the RMD QoS model
- NF edge - a QoS NF that is located at the boundary of an

administrative domain, e.g., Diffserv.

- NF egress - an edge QoS NF that handles the traffic as it leaves the domain.
- NF ingress - an edge QoS NF that handles the traffic as it enters the domain.
- NF interior - a QoS NF that is part of an administrative domain, e.g., Diffserv, and is not an NF edge.
- NTLP stateful node - a NTLP aware node that maintains a NTLP transport layer state.
- NTLP stateless node - a NTLP aware node that does not maintain a NTLP transport layer state.
- Stateful QNE - a QNE that supports the QoS NSLP stateful operation.
- Stateless QNE - a QNE that supports the QoS NSLP stateless operation.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

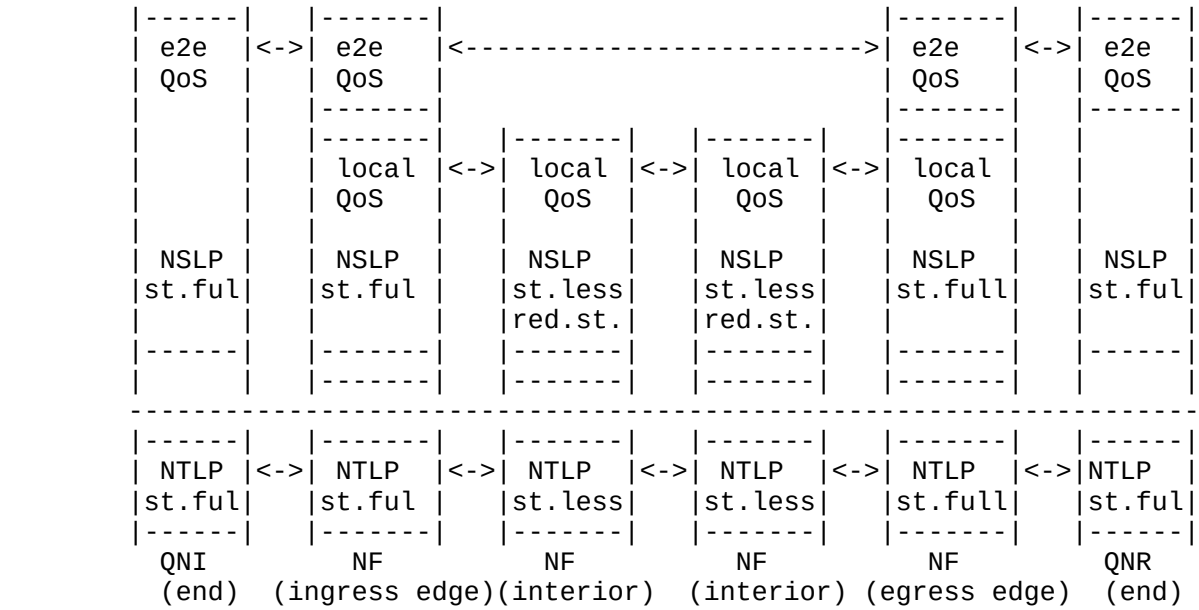
2. Protocol model

The protocol model of the scalable QoS model is shown in Figure 1. Consider an end-to-end QoS model, in which NF nodes between End and Edge nodes install and maintain per-flow QoS-NSLP states. QoS-NSLP messages are transported by NTLP, therefore these NF-s are NTLP stateful as well. In the Edge node of the Diffserv domain end-to-end QoS-NSLP messages generate local QoS-NSLP messages. The original QoS-NSLP messages are handed over to the next NTLP stateful node, e.g. to the egress edge node.

The local QoS-NSLP messages are transported by the NTLP datagram mode (IP or UDP/IP protocols) to egress edge. The RMD QoS model is

suitable for a reduced state or stateless form of operation. When processed by interior (stateless) nodes the QoS NSLP use options to store minimum number of states. Some state, e.g. per class, for the QoS model related data may be held at these NF interior nodes. The QoS NSLP also requests that the NTLP use different transport characteristics (i.e. sending of messages in datagram mode, and not retaining optional path state, i.e., NTLP stateless mode).

In the edge node, the QSpec of the end-to-end QoS-model is transformed into resource units. These resources are requested dynamically per PHB group and in all nodes in the communication path from an ingress edge node to egress edge node.



st.ful : statefull
st.less : stateless
st.less red.st. : stateless or reduced state

Figure 1 Protocol model of stateless/reduced state operation

In case of measurement based method a local RESERVE message is sent to check the availability of resources before flows are admitted. In the interior nodes two QoS-NSLP states per PHR group are installed, which do not have to be maintained (by refresh) during the

connection. One state stores the measured user traffic load associated to PHB and another state stores the maximum traffic load per PHB group that can be admitted.

In case of reservation-based method per PHB group aggregated reservations states are installed and these are maintained by sending RESERVE messages. The reservation-based PHR installs and maintains one reservation state per PHB, i.e., per DSCP, in all the nodes located in the communication path from the NF ingress node up to the NF egress node. The reservation states can be represented by constant number of parameter set. This state represents the number of currently reserved resource units that are carried by the PHR object for the admitted incoming flows. Thus, the NF ingress node generates for each incoming flow a PHR Object that is included into a local RESERVE(RMD-QSPEC), signaling only the resource units requested by this particular flow. These resource units if admitted is added to the currently reserved resources per PHB.

For each PHB a threshold is maintained that specifies the maximum number of resource units that can be reserved. This threshold could, for example, be statically configured.

The per-PHB group reservation states can be created and maintained by the combination of reservation soft state and explicit release principles. When the reservation soft state principle is used, a finite lifetime is set for the length of the reservation. These reservation states are refreshed by sending periodic refresh messages. The reserved resources for a particular flow can also be explicitly released from a PHB reservation state by means of a PHR release message. The usage of explicit release enables the instantaneous release of the resources regardless of the length of the refresh period. This allows a longer refresh period, which also reduces the number of periodic refresh messages.

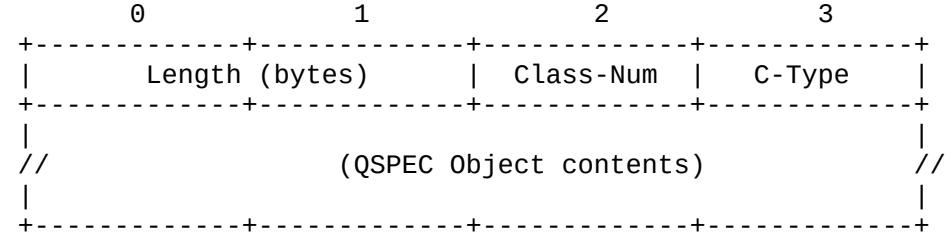
NTLP stateless QNEs are not able to perform message bundling, message fragmentation and reassembly (in the NTLP) or congestion control. They are not able to establish and maintain security associations with their neighbors, which means, they can only be applied in a trusted environment.

3. Message format

The format of the messages used by the RMD QoS model complies with the QoS-NSLP specification. As specified in [QoS-NSLP], for each QoS-NSLP message type, there is a set of rules for the permissible choice of object types. These rules are specified using Backus-Naur Form (BNF) augmented with square brackets surrounding optional sub-sequences. The BNF implies an order for the objects in a message. However, in many (but not all) cases, object order makes no logical difference. An implementation should create messages with the objects in the order shown here, but accept the objects in any permissible order.

In general, an NSIS QoS model specifies the QoS-NSLP QSPEC object. QSPEC object contains QoS model-specific Control Information components, denoted as LCI in this draft. Using these fields QoS-model defines which objects of QoS-NSLP are used and it also specifies the QoS-model specific rules.

As specified in the QoS-NSLP draft, the QSPEC object consists of one or more 32-bit words with a one-word header, with the following format:



An object header has the following fields:

Length

A 16-bit field containing the total object length in bytes. Must always be a multiple of 4, and at least 4.

Class-Num

Identifies the object class; For a QSPEC object this value

is set to 8.

C-Type

Object type, unique within Class-Num. For a QSPEC object is specifying the QoS-model ID. For the time being this value is for the RMD-QoS model chosen to be 10.

The RMD QSPEC object contains two objects, the per hop reservation (PHR) and the per domain reservation (PDR) objects. The PHR object is used (and processed by all QoS-NSLP nodes in the edge-to-edge domain, i.e., NF (edge nodes) and NF (interior nodes). The PDR object is only used (processed) by the NF (edge nodes). The PHR object contains the local QSPEC object for intra-domain communication and reservation. The PDR object contains additional information that is needed by the QNI (edge nodes) and is not available (carried) by the PHR object. The definition of the RMD-QSPEC object is given in Section 7.

The format of a local Reserve (RMD-QSPEC) message used by the RMD QoS model is as follows:

```
<Reserve Message> ::= <Common Header>

                      <RSN> [ <SCOPING> ]  <RESPONSE_REQUEST>

                      <TIME_VALUES> [ <SESSION ID> ]

                      <POLICY_DATA> <PHR> [ <PDR> ]
```

The format of a local Query (RMD-QSPEC) message used by the RMD QoS model is as follows:

```
<Query Message> ::= <Common Header>

                        [ <SCOPING> ] <RESPONSE_REQUEST>

                        [ <TIME_VALUES> ] [ <SESSION ID> ]

                        <POLICY_DATA> <PHR> [ <PDR> ]
```

The format of a local Response (RMD-QSPEC) message used by the RMD

QoS model is as follows:

<Response Message> ::= <Common Header>

<SCOPING> [<ERROR_SPEC>]

<PDR>

The format of an end-to-end Response (PDR) message that is used by the RMD QoS model to carry a PDR object is as follows:

<Response Message> ::= <Common Header>

[<RSN>] [<SCOPING>] [<ERROR_SPEC>]

<PDR> [<QSPEC> ..]

The format of a local Notify (RMD-QSPEC) message used by the RMD QoS model is as follows:

<Notify Message> ::= <Common Header>

[<ERROR_SPEC>] <PDR>

All objects, except the RMD QSPEC objects, are specified in [QoS-NSLP]. All local messages used by the RMD QoS model must operate in the NTLP Datagram mode (see [GIMPS]). Therefore, the NSLP functionality available in all QoS NSLP nodes that are able to support the RMD QoS model must require from the local NTLP functionality available in these nodes to operate in the Datagram mode. The QoS NSLP may want to restrict the handling of its messages to specific nodes. This functionality is needed to support layering, when only the edge QNEs (e.g., NF edges) of a domain process the message. This requires a mechanism at the QoS NSLP level to bypass intermediates nodes between the edges of the domain.

As a suggestion, the QoS-NSLP draft [QoS-NSLP] identifies two ways for bypassing intermediate nodes, e.g., NF interior nodes. One solution is for the end-to-end session to carry a different protocol

ID (QoS-NSLP-E2E-IGNORE protocol ID, similar to the RSVP-E2E-IGNORE that is used for RSVP aggregation ([RFC3175]). Another solution is based on the use of multiple levels of the router alert option. In that case, internal routers, e.g., NF interior nodes, are configured to handle only certain levels of router alerts. The choice between both approaches or another approach that fulfills the requirement is left to the NTLP design.

4. Normal operation for unidirectional reservations

4.1. RMD specific new reservation: successful operation

The QNI performs generates the initial RESERVE message, and it is forwarded by the NTLP as usual [GIMPS]. At the QNEs at the edges of the RMD domain the processing is different and the nodes support two QoS models.

At the ingress the original RESERVE message is forwarded but using facilities provided by the NTLP to bypass the stateless or reduced-state nodes, see Figure 2. After the initial discovery phase using datagram mode, connection mode between the ingress and egress can be used. At the egress node the RESERVE message is then forwarded normally.

The NF ingress has to request NTLP to activate the QoS-NSLP-E2E-IGNORE feature (its specification depends on NTLP and QoS-NSLP standardization) for transporting end-to-end QoS-NSLP messages. In this way all the NF interior nodes ignore the processing of the end-to-end RESERVE message. The resource description (QSPEC) used by the end-to-end QoS model is transformed into RMD traffic class (PHR) resource units (needed for the local QSPEC).

In order to make a RMD query or a RMD reservation a local RESERVE(RMD-QSPEC) message is generated by the NF ingress edge. At the ingress a second RESERVE message (i.e., local RESERVE(RMD-QSPEC)) is also built. This makes use of a QoS model suitable for a reduced state or stateless form of operation (such as the RMD per hop reservation).

Before generating this message, the RMD QoS model functionality is using the RMD traffic class (PHR) resource units for admission

control.

* In case of the RMD reservation-based procedure, these resources, if admitted are added to the currently reserved resources per traffic class (PHB) and, therefore, they become a part of the per RMD traffic class (PHB) reservation state. Furthermore, the value of the PHR_TTL field in the <PHR> object has to be set to one. The PHR_TTL value is used to count the number of RMD NSIS aware nodes that successfully processed the reservation based <PHR> object.

* in case of the RMD measurement based method, if these resources are admitted, using a MBAC algorithm, the number of this resources will be used to update the MBAC algorithm.

The session ID used by this message must be associated to a DSCP value. The IP destination address of this message must be the same as the IP destination address of the end-to-end RESERVE message. If the end-to-end RESERVE request is satisfied locally, the NF ingress node generates a reservation request <PHR> object denoted as "PHR_Resource_Request" and it may generate a reservation request PDR object denoted as "PDR_Reservation_Request", (see Section 8). These two objects form the RMD-QSPEC object. This local RESERVE (RMD-QSPEC) message must include a <PHR>, (i.e., PHR_Resource_Request) object, and it may include a <PDR>, (i.e., PDR_Reservation_Request) object. The non-default values of the objects contained in this local RESERVE message must be set by the NF ingress edge as follows:

* the value of the <RSN> object should be the same as the value of the RSN object of the end-to-end RESERVE message.

* the SCOPING object must not be included into the message, meaning that a default scoping of the message is used. Therefore, the NF edges must be configured as boundary nodes and the NF interior nodes should be configured as interior (intermediary) nodes.

* The value of the <RESPONSE_REQUEST> object must contain the Response Identification Information (RII) that is unique within a session and different for each message (see [(QoS-NSLP]). In general downstream nodes that desire responses may keep track of this RII to identify the RESPONSE when it passes back through them.

* the value of the <TIME_VALUES> object must be calculated and set by the NF ingress node.

* the value of the <SESSION_ID> object must be the session ID associated to the end-to-end RESERVE message.

* the value of the <POLICY_DATA> object depends on the used policy;

* the PHR resource units must be included into the "Requested Resources" field of the PHR object;

* the value of the C field of <PHR> object is set to 1 (PHR_Resource_Request)

* the value of the PHR_TTL field in the <PHR> object has to be set to one. The PHR_TTL value is used to count the number of RMD reservation based NSIS aware nodes that successfully processed the reservation based <PHR> object.

* the <PDR> object may not be included into the message.

The RMD query procedure is needed in the case of the RMD measurement based method, see e.g., [RIMA], while the RMD reservation procedure is needed in case of reservation-based method, see e.g., [RODA].

When processed by NF interior (stateless) nodes the QoS NSLP processing excercises its options to not keep state wherever possible, so that no QoS NSLP state is stored. Some state, e.g. per traffic class, for the RMD QoS model related data may be held at these interior nodes. The QoS NSLP also requests that the NTLP use different transport characteristics (i.e. sending of messages in datagram mode, and not retaining optional path state).

Nodes, such as those in the interior of the stateless or reduced-state domain, that do not retain reservation state (and so cannot use summary refreshes) cannot send back RESPONSE messages.

The non-default values of the objects contained in the local RESERVE (RMD- QSPEC) message must be set by each NF interior node as follows:

* the values of the <RSN>, [<SCOPING>], <RESPONSE_REQUEST>, <TIME_VALUES>, <SESSION_ID>, <POLICY_DATA> objects are not changed, i.e., equal to the values set by the NF ingress edge;

* the value of "Resource Request" field of the <PHR> object is used by the NF interior node for admission control.

* In case of the RMD reservation-based procedure, if these resources are admitted are going to be added to the currently reserved resources per PHB and therefore they will become a part of the per RMD traffic class (PHB) reservation state. Furthermore, the value of the PHR_TTL field in the <PHR> object has to be increased by one. The PHR_TTL value is used to count the number of RMD reservation based NSIS aware nodes that successfully processed the reservation based <PHR> object.

* in case of the RMD measurement based method, if these resources are admitted, using a MBAC algorithm, the number of this resources will be used to update the MBAC algorithm.

When the local RESERVE (RMD-QSPEC) is received by the NF egress node a binding of the session associated with the local RESERVE (RMD-QSPEC) (the DSCP session) with the session included in its SESSION_ID object must be accomplished. The session included in the <SESSION_ID> object is the session associated with the end-to-end RESERVE.

The <PHR> object (and if available the <PDR> object) are read and processed by the RMD QoS mode functionality. The value of "Resource Request" field of the <PHR> object is used by the NF egress node for traffic class admission control.

* In case of the RMD reservation-based procedure, if these resources are admitted are going to be added to the currently reserved resources per PHB and therefore they will become a part of the per PHB reservation state. Furthermore, the value of the PHR_TTL field in the <PHR> object has to be increased by one. The PHR_TTL value is used to count the number of RMD reservation based NSIS aware nodes that successfully processed the reservation based <PHR> object.

* In case of the RMD measurement based method, if these resources are admitted, using a MBAC algorithm, the number of these resources are used to update the MBAC algorithm.

At the NF egress node the local RESERVE (RMD-QSPEC) message is

interpreted in conjunction with the reservation state from the end-to-end RESERVE message (using information carried in the message to correlate the signalling flows, i.e., SESSION_ID). The end to end RESERVE message is only forwarded further if the processing of the local RESERVE (RMD-QSPEC) message was successful at all nodes in the RMD domain, otherwise the end-to-end reservation is regarded as having failed to be installed.

Since NTLP neighbour relations are not maintained in the reduced-state or stateless RMD domain, only sender initiated signalling can be supported. If a bi-directional reservation is required then the interior QoS model must provide an object that requests the egress node to generate a sender initiated session in the reverse direction. The NF egress nodes should de-activate this NTLP QoS-NSLP-E2E-IGNORE feature.

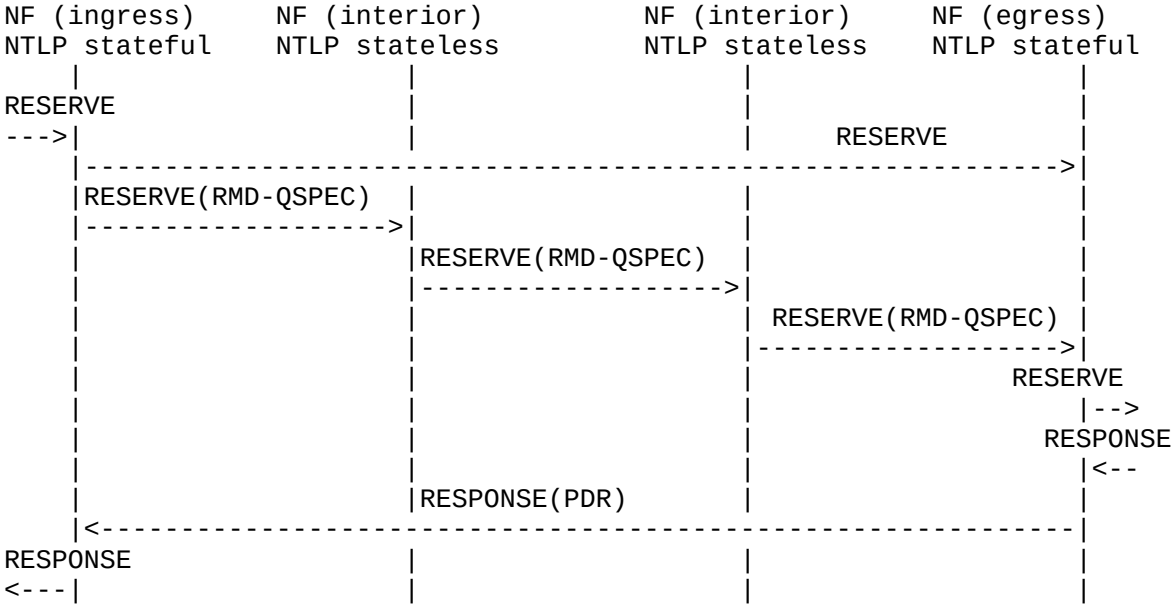


Figure 2: Basic operation of successful reservation procedure used by the RMD QoS model

After a positive response (successfully processed end-to-end REPORT message) message arrives, in return, at the NF egress edge, the NF egress node must include a PDR object (i.e., PDR_Reservation_Report)

into this end-to-end RESPONSE message (see Section 3). NF egress asks NTLP to activate the QoS-NSLP-E2E-IGNORE feature. In this way all the NF interior nodes must ignore the processing of the end-to- end RESPONSE message. The REPORT message is sent to its upstream QoS-NSLP neighbor. Note that this message will use a NTLP connection mode.

The non default values of the objects contained in the end-to-end RESPONSE message must be set by the NF egress node as follows:

* the values of the [<RSN>], [<SCOPING>], [<ERROR_SPEC>], [QSPEC ..] objects are set by the standard QoS-NSLP protocol functions;

* the value of the MsgType field of the PDR object sould be set to 4 (PDR_Reservation_Report).

* The value of the F-Type depends on the policy used by the NF egress node;

* The value of the EP-Type field of the PDR message should be equal to the QoS-NSLP protocol ID.

* The value of the Flow ID field of the PDR object depends on the policy used by the NF egress node.

This RESPONSE message is received by the NF ingress node. The non default values of the objects contained in the end-to-end RESPONSE message must be set by the NF ingress node as follows:

* the values of the [<RSN>], [<SCOPING>], [<ERROR_SPEC>], [QSPEC ..] objects are set by the standard QoS-NSLP protocol functions;

* the PDR object has to be processed and removed by the RMD QoS model functionality in the NF ingress node. The RMD QoS model functionality is notified by reading the "M" filed of the <PDR> object that the reservation has been successful.

Future versions of this draft will include more details on the RMD successful reservation procedure.

4.2. RMD specific new reservation: unsuccessful operation

The NF ingress and the NF interior and NF egress nodes processes and forwards the end-to-end RESERVE message and the local RESERVE (RMD-QSPEC) message in the same way as specified in Section 4.1. The main difference between the unsuccessful operation and successful operation is that one of the NF nodes will not admit the request due to lack of resources. This also means that the NF edge node does not forward the end-to-end RESERVE message towards the QNR node, but it is discarded.

When an end-to-end RESERVE message arrives to the NF ingress edge and if there are no resources available locally, the NF ingress node rejects this end-to-end RESERVE message and send a REPORT message back to the sender in a standard QoS-NSLP method.

Any NF edge or NF interior node that receives a "PHR_Resource_Request" <PHR> object it must identify the traffic class state (DSCP).

In case of the RMD reservation based scenario, if the reservation request, i.e., <PHR> object, is not admitted by the NF node then the T and M fields of the PHR object have to be set to 1. In this case the PHR_TTL counter value must not be increased.

In case of the RMD measurement based scenario, if the reservation request, <PHR> object, is not admitted by the MBAC algorithm used at the NF node, then the M field of the PHR object have to be set to 1.

In general if a NF interior node receives a <PHR> object, of type PHR_Resource_Request, with the M field of the <PHR> object set to "1" then this <PHR> object must not be processed, i.e., its fields will not be read and/or modified.

In both scenarios, i.e., RMD reservation based and RMD measurement based scenario, when the "M" marked local RESERVE (RMD-QSPEC) is received by the NF egress node (see Figure 3) a binding of the session associated with the local RESERVE (RMD-QSPEC) (the DSCP session) with the session included in its SESSION_ID object must be accomplished. The session included in the <SESSION_ID> object is the session associated with the end-to-end RESERVE.

The NF egress node must generate a REPORT message that will have to be sent to its previous stateful QoS-NSLP hop. This message must

include a <PDR> object (of type PDR_Reservation_Report). Note that this message will use a NTLP connection mode.

The non-default values of the objects contained in the end-to-end RESPONSE message must be set by the NF egress node as follows:

* the values of the [<RSN>], [<SCOPING>], [<ERROR_SPEC>], [QSPEC ..] objects are set by the standard QoS-NSLP protocol functions.

* the value of the MsgType field of the PDR object sould be set to 4 (PDR_Reservation_Report).

* The value of the PHR_TTL value of the <PHR> object included in the received "M" marked local RESERVE (RMD-QSPEC) message must be included in the PDR_TTL field of the <PDR> object.

* The value of the "M" field of the <PDR> object must be set to 1.

* The value of the F-Type depends on the policy used by the NF egress node;

* The value of the EP-Type field of the PDR message should be equal to the QoS-NSLP protocol ID.

* The value of the Flow ID field of the PDR object depends on the policy used by the NF egress node.

The non-default values of the objects contained in the end-to-end RESPONSE (PDR) message must be set by the NF ingress node, which receives this message, as follows:

* the values of the [<RSN>], [<SCOPING>], [<ERROR_SPEC>], [QSPEC ..] objects are set by the standard QoS-NSLP protocol functions;

* the PDR object has to be processed and removed by the RMD QoS model functionality in the NF ingress node. The RMD QoS model functionality is notified by reading the "M" field of the <PDR> object that the reservation has been unsuccessful. In case of a

RMD reservation based scenario, the RMD QoS model functionality, has to start an RMD specific release procedure (see Section 4.5.2).

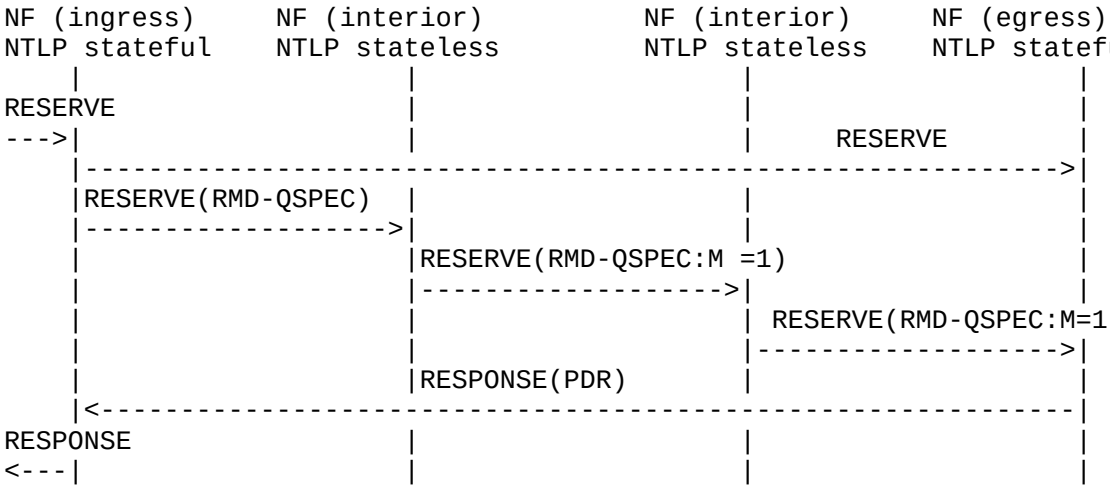


Figure 3: Basic operation during unsuccessful reservation initiation used by the RMD QoS model

Future versions of this draft will include more details on the RMD unsuccessful reservation procedure.

4.3. RMD specific refresh reservation

In case of RMD measurement-based method, QoS-NSLP states in the RMD domain are not maintained, therefore, the end-to-end RESERVE (refresh) message is sent directly to NF egress edge.

The refresh procedure in case of RMD reservation-based method follows a similar scheme as the reservation process, shown in Figure 2. Until reservations messages arrive within the time-out period, the corresponding number of resource units is not removed. However, in this scenario the generation of the end-to-end RESERVE message, by NF edges is generated and sent to the next hop, depends on the maintained refreshed period (see [QoS- NSLP]). In other words the moment that the end-to-end refresh RESERVE message is sent by the NF

egress node downstream to its next hop, depends on the maintained refresh period and not on the moment that the local RERSERVE (RMD-QSPEC) message, which is bound to it, is received by the NF egress node. QoS-NSLP-E2E-IGNORE feature of NTLP must be activated by NF ingress and deactivated by the NF egress node.

The RMD QoS model functionality available in the NF ingress node must be able to generate local RESERVE (RMD-QSPEC) messages that will be sent towards the NF egress node, in order to refresh the RMD traffic class state in the NF edges and interior nodes. Before generating this message, the RMD QoS model functionality is using the RMD traffic class (PHR) resource units for refreshing the RMD traffic class state.

Note that the RMD traffic class refresh periods should be equal in all NF edge and NF interior nodes and should be smaller (default: more than two times) than the refresh period at the NF ingress node used by the end-to-end RESERVE message. This local RESERVE (RMD-QSPEC) message must include a <PHR>, (i.e., PHR_Refresh_Update) object, and it may include a <PDR>, (i.e., PDR_Refresh_Request) object.

The selection of the IP source and destination address of this message depends on if and how the different end-to-end flows can be aggregated by the NF ingress node. Note that this aggregation procedure is different than the RMD traffic class aggregation procedure. One example approach is the approach used by the RSVP aggregation scenario ([RFC3175]), where the IP source address of this message is the IP address of the aggregator (i.e., NF ingress edge) and the IP destination address of this message is the IP address of the De-aggregator (i.e., NF egress). Another example approach is the approach used in "RSVP Refresh Overhead Reduction Extensions" ([RFC2961]). If no aggregation procedure is possible then the IP destination address of this message should be equal to the IP destination address of its associated end-to-end RESERVE message.

An example of this RMD specific refresh operation can be seen in Figure 4.

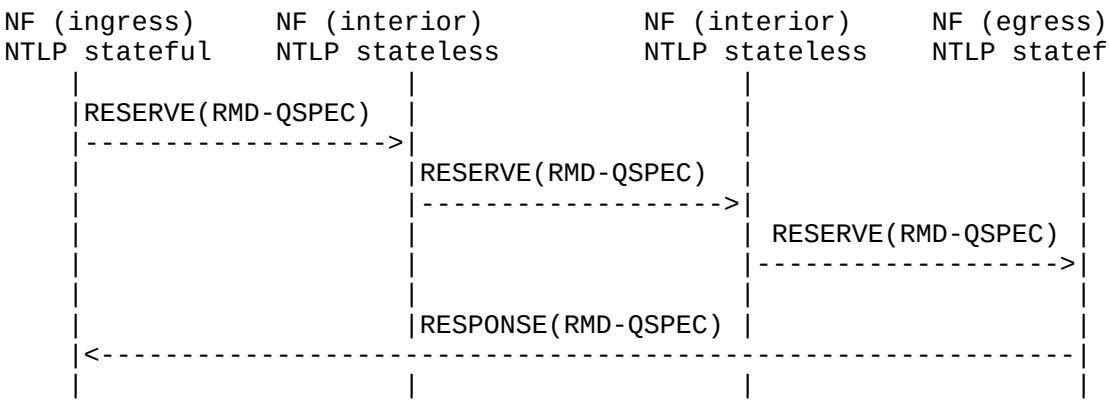


Figure 4: Basic operation of RMD specific refresh procedure

The most of the non default values of the objects contained in this message are set by the NF ingress in the same way as described in Section 4.1. The following objects are set differently:

* the <SESSION_ID> object is not included in this message.

* the PHR resource units must be included into the "Requested Resources" field of the <PHR> object. The value of the "Requested Resources" depends on if and how the different end-to-end flows can be aggregated by the NF ingress node (e.g., the sum of all the PHR requested resources of the aggregated flows). If no flow aggregation is accomplished by the NF ingress node, then the value of the "Requested Resources" field should be equal to the "Requested Resources" field of its associated new (initial) local RESERVE (RMD-QSPEC) message;

* the value of the C field of <PHR> object is set to 2 (PHR_Refresh_Update)

* the value of the <PDR> object may not be included into the message.

The local RESERVE (RMD-QSPEC) message is received and processed by the NF interior nodes. Any NF edge or NF interior node that receives a "PHR_Refresh_Update" <PHR> object it must identify the traffic class state (DSCP). The most of the non default values of the objects contained in this refresh local RESERVE (RMD- QSPEC) message are set by

a NF interior node in the same way as described in Section 4.1.

The following objects are set and processed differently:

* the value of "Resource Request" field of the <PHR> object is used by the NF interior node for refreshing the RMD traffic class state. If the refresh procedure cannot be fulfilled then the M field of the <PHR> object has to be set to 1.

Any NF edge or NF interior node that receives a "PHR_Resource_Release" <PHR> object it must identify the traffic class state (DSCP) and release the requested resources included in the "Requested Resources" field.

Any <PHR> object of type PHR_Refresh_Update, whether it is marked or not, is always processed (the "Requested Resources" field), but marked bits are not changed.

The local RESERVE (RMD-QSPEC) message is received and processed by the NF egress node. The <PHR> object (and if available the <PDR> object) are read and processed by the RMD QoS mode functionality. The value of "Resource Request" field of the <PHR> object is used by the NF egress node for refreshing the RMD traffic class state. If the refresh procedure cannot be fulfilled then the M field of the <PHR> object has to be set to 1.

A new local RESPONSE (RMD-QSPEC) message is generated by the NF egress node. This message must include a <PDR> object (of type PDR_Refresh_Report).

This message must be sent to the NF ingress node, i.e., previous stateful hop. This can for example be accomplished by using the value of the REQUEST_RESPONSE object (RII) included in the received local RESERVE (RMD- QSPEC) message. In general downstream nodes that desire responses may keep track of this RII to identify the RESPONSE when it passes back through them. This RII value will be included in the <SCOPING> object of the generated local RESPONSE (RMD-QSPEC) message. The most of the non default values of the objects contained in this refresh local RESPONSE (RMD-QSPEC) message are set by a NF egress node in the same way as described in Section 4.1.

The following objects are set and processed differently:

* the value of the <SCOPING> object is equal to the RII that is used by the NF ingress to identify the RESPONSE when it passes back through it. This value was carried by the local RESERVE (RMD-QSPEC) message in the <RESPONSE_REQUEST> object.

* MsgType field of the PDR object should be set to 5 (PDR_Refresh_Report).

* The value of the "M" field of the <PDR> object must be equal to the value of the "M" field of the <PHR> object that was carried by its associated local RESERVE (RMD-QSPEC) message.

When the local RESPONSE (RMD-QSPEC) message is received by the NF ingress node, then:

* the values of the [<RSN>], [<SCOPING>], [<ERROR_SPEC>], [QSPEC ..] objects are processed by the standard QoS-NSLP protocol functions;

* the PDR object has to be processed and removed by the RMD QoS model functionality in the NF ingress node. The RMD QoS model functionality is notified by reading the "M" field of the <PDR> object that the refresh procedure has been successful or unsuccessful. All session(s) (in case of the flow aggregation procedure there will be more than one sessions) associated with this RMD specific refresh session must be informed about the success or failure of the refresh procedure. In case of failure, the NF ingress node has to generate (in a standard QoS-NSLP way) an error end-to-end REPORT message that will be sent towards QNI.

Future versions of this draft will include more details on the RMD specific refresh reservation procedure.

4.4. RMD specific modification of reservation

When the RMD QoS model functionality of the NF ingress node receives an end- to-end RESERVE message that is requesting a modification on the number of reserved resources then the following process can be realized. When the modification request requires an increase on the number of reserved resources, then the RMD QoS model functionality of the ingress node will have to subtract the old and already reserved

number of resources from the number of resources included in the new modification request. The result of this subtraction should be introduced within a PHR_Resource_Request <PHR> object as the "Requested Resources" value. If a NF edge or NF interior node is not able to reserve the number of requested resources, then the "PHR_Resource_Request" <PHR> object will be marked. In this situation the RMD specific operation for a unsuccessful reservation functionality will be applied for this case (see Section 4.2).

When the modification request requires a decrease on the number of reserved resources, then the NF ingress node will have to subtract the number of resources included in the new modification request from the old and already reserved number of resources. The result of this subtraction should be introduced in a PHR_Release_Request <PHR> object, and a RMD specific release procedure should be accomplished (see Section 4.5.2) Future versions of this draft will include more details on the RMD specific modification reservation procedure.

4.5. RMD specific release procedure

Resources in NF interior nodes are removed after time-out if refresh message does not arrive in time in case of reservation base method. This soft state behavior provides certain robustness for the system ensuring that unused resources are not reserved for long time. However, if even more efficient resource management is needed, resources can be removed by explicit release procedure within the refresh period.

In general, when the RMD QoS model functionality of a NF edge or NF interior node processes a "PHR_Release_Request" <PHR> object it must identify the DSCP and estimate the refresh period where it last signalled the resource usage (where it last processed a "PHR_Refresh_Update" <PHR> object). This may be done by, for example, giving the opportunity to an NF ingress node to calculate the time lag, say T_lag, between the last sent "PHR_Refresh_Update" <PHR> object and the "PHR_Release_Request" <PHR> object. The value of this time lag (T_lag), is first normalized to the length of the refresh period, say T_period. In other words the ratio between this time lag, T_lag, and the length of the refresh period, T_period, is calculated. This ratio is then introduced into the "Delta T" field of the "PHR_Release_Request" <PHR> object. When a node (NF edge or NF interior) receives this "PHR_Release_Request" <PHR> object, it will

have to store its arrival time. Then it will calculate the time difference, say Tdiff, between this arrival time and the start of the current refresh period, T_period. Furthermore, this node will have to derive the value of the time lag, T_lag, from the "Delta T" field. This can be found by multiplying the value included in the "Delta T" field with the length of the refresh period, T_period. If the derived time lag, T_lag, is smaller than the calculated time difference, T_diff, then this node MUST decrease the PHB reservation state with the number of resource units indicated in the "Requested Resources" field of the "PHR_Release_Request" message, but not below zero.

An RMD specific release procedure can be triggered by an end-to-end RESERVE with a TEAR flag set ON (see Section 4.5.1) or it can be triggered by either a RESPONSE or NOTIFY message that includes a marked (i.e., "M" and/or "S" field is 1) <PDR> object of PDR_Reservation_Report type (see Section 4.2) or PDR_Congestion_Report type (see Section 6.2).

4.5.1. Triggered by a RESERVE message

This RMD explicit release procedure can be triggered by a tear (TEAR flag set ON) end-to-end RESERVATION message. When a tear (TEAR flag set ON) end-to-end RESERVE message arrives to the NF ingress edge then the NF ingress node will have to process the message in a standard QoS-NSLP way (see [QoS-NSLP]). In addition to this the RMD QoS model functionality must be notified. The RMD QoS model functionality will generate a local RESERVE (RMD-QSPEC) message. Before generating this message, the RMD QoS model functionality is using the RMD traffic class (PHR) resource units for a RMD release procedure. This can be achieved by subtracting the amount of RMD traffic class requested resources from the total reserved amount of resources stored in the RMD traffic class state.

This local RESERVE (RMD-QSPEC) message must include a <PHR>, (i.e., PHR_Resource_Release) object and it may include a <PDR>, (i.e., PDR_Release_Request) object. An example of this operation can be seen in Figure 5.

The most of the non default values of the objects contained in the tear local RESERVE message are set by the NF ingress node in the same way as described in Section 4.1.

The following objects are set differently:

* The <RESPONSE_REQUEST> object is not included in this message. This is because the NF ingress node does not need to receive a response from the NF egress node.

* The TEAR flag is set to ON (T = 1);

* the PHR resource units must be included into the "Requested Resources" field of the PHR object;

* the value of the PHR_TTL field in the <PHR> object has to be set to one. The PHR_TTL value is used to count the number of RMD reservation based NSIS aware nodes that successfully processed the reservation based <PHR> object.

* the value of the Delta_T field of the <PHR> is calculated by the RMD QoS model functionality (see introductory part of Section 4.5)

* the value of the C field of <PHR> object is set to 3 (PHR_Resource_Release)

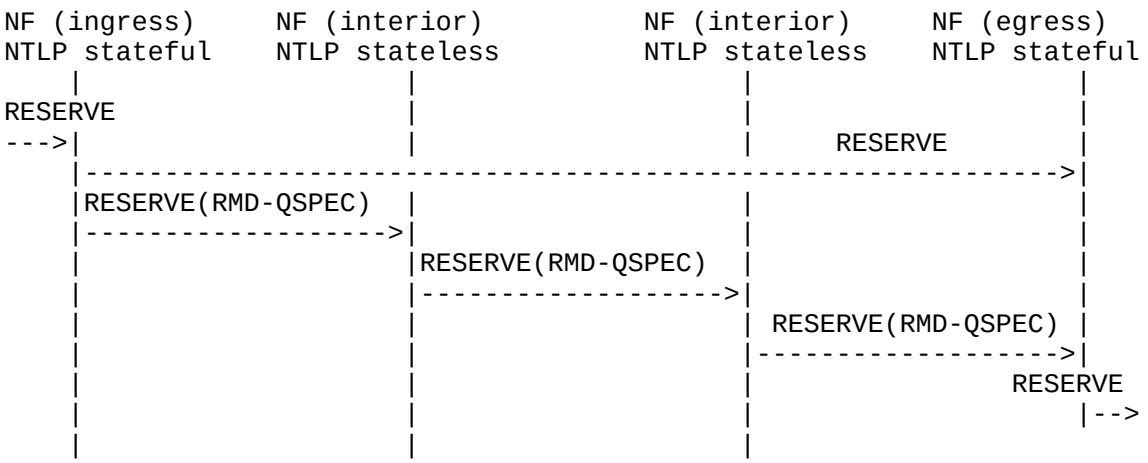


Figure 5: Explicit release triggered by RESERVE used by the RMD QoS model

The local tear RESERVE (RMD-QSPEC) message is received and processed by the NF interior nodes. The most of the non default values of the objects contained in this refresh local RESERVE (RMD-QSPEC) message are set by a NF interior node in the same way as described in Section 4.1. The following objects are set and processed differently:

* Any NF interior node that receives a "PHR_Resource_Release" <PHR> object it must identify the traffic class state (DSCP) and release the requested resources included in the "Requested Resources" field. This can be achieved by subtracting the amount of RMD traffic class requested resources, included in the "Requested Resources" field, from the total reserved amount of resources stored in the RMD traffic class state. The value of the Delta_T field of the <PHR> object is used during the release procedure as explained in the introductory part of Section 4.5.

The local tear RESERVE (RMD-QSPEC) message is received and processed by the NF egress node. The <PHR> object (and if available the <PDR> object) are read and processed by the RMD QoS mode functionality. The value of "Resource Request" field of the <PHR> object and the value of the "Delta_T" field of the <PHR> object must be used by the RMD release procedure. This can be achieved by subtracting the amount of RMD traffic class requested resources, included in the "Requested Resources" field, from the total reserved amount of resources stored in the RMD traffic class state.

The end-to-end RESERVE message is forwarded by the next hop (i.e., NF egress) only if the local tear RESERVE (RMD-QSPEC) message arrives at the NF egress node. The QoS-NSLP-E2E-IGNORE feature of NTLP must be deactivated.

Future versions of this draft will include more details on this RMD specific release procedure.

4.5.2. Triggered by a marked RESPONSE or NOTIFY message

This RMD explicit release procedure can be triggered by either a RESPONSE message with a "M" marked <PDR> object (see Section 4.2) or a NOTIFY message (see Section "Severe congestion") with a "M" or "S" marked <PDR> object. This RMD specific release procedure can be terminated at any NF interior node or NF edge node. The RMD specific

explicit release procedure that is terminated at a NF interior (or NF edge) node is denoted as RMD specific partial release procedure. This explicit release procedure can be, for example, used during a RMD specific operation for unsuccessful reservation (see Section 4.2) or severe congestion (see Section 6.2). When the RMD QoS model functionality of a NF ingress node receives a "M" or "S" marked <PDR> object of type PDR_Reservation_Report or PDR_Congestion_Report, it must start a RMD partial release procedure. The NF ingress node generates a local RESERVE (RMD-QSPEC) message. Before generating this message, the RMD QoS model functionality is using the RMD traffic class (PHR) resource units for a RMD release procedure. This can be achieved by subtracting the amount of RMD traffic class requested resources from the total reserved amount of resources stored in the RMD traffic class state.

When the generation of the local RESERVE (RMD-QSPEC) message is triggered by a RESPONSE (PDR) message then the this local RESERVE (RMD-QSPEC) message must include a <PHR>, (i.e., PHR_Resource_Release) object and a <PDR>, (i.e., PDR_Release_Request) object. An example of this operation can be seen in Figure 6.

When the generation of the local RESERVE (RMD-QSPEC) message is triggered by a NOTIFY (PDR) message then the this local RESERVE (RMD-QSPEC) message must include a <PHR>, (i.e., PHR_Resource_Release) object and it may include a <PDR>, (i.e., PDR_Release_Request) object. An example of this operation can be seen in Figure 6.

The most of the non default values of the objects contained in the tear local RESERVE (RMD-QSPEC) message are set by the NF ingress node in the same way as described in Section 4.5.1.

The following objects are set differently:

* The value of the "M" field of the <PHR> object must be set to 1.

* When the tear local RESERVE message is triggered by a NOTIFY message, then the value of the "S" field of the <PHR> object must be set to "1".

* When the generation of the local RESERVE (RMD-QSPEC) message is triggered by a NOTIFY (PDR) message then the this local RESERVE (RMD-QSPEC) message does not include a <PDR>.

* When the tear local RESERVE message is triggered by a RESPONSE message, then the value of the PDR_TTL value of the <PDR> object included in the

received "M" marked local RESPONSE (PDR) message must be included in the PDR_TTL field of the <PDR> object. The value of the F-Type depends on the policy used by the NF egress node. The value of the EP-Type field of the PDR message should be equal to the QoS-NSLP protocol ID. The value of the Flow ID field of the PDR object depends on the policy used by the NF egress node.

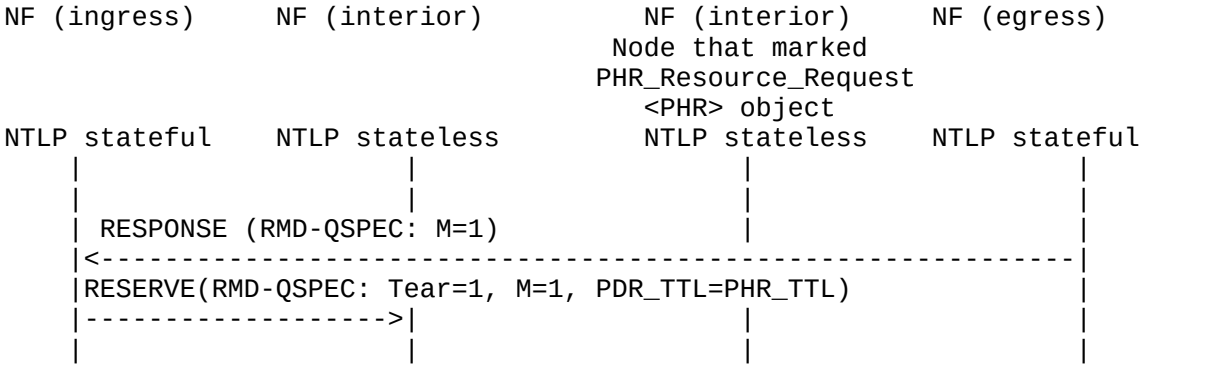


Figure 6: Basic operation during RMD explicit release procedure triggered by RESERVE used by the RMD QoS model

Any NF edge or NF interior node that receives a "PHR_Resource_Release" <PHR> object it must identify the traffic class state (DSCP) release the requested resources included in the "Requested Resources" field. This can be achieved by subtracting the amount of RMD traffic class requested resources, included in the "Requested Resources" field, from the total reserved amount of resources stored in the RMD traffic class state. The value of the Delta_T field of the <PHR> object is used during the release procedure as explained in the introductory part of Section 4.5. Furthermore, the PHR_TTL value included in the <PHR> object is increased by one. If the value of "M" field of the PHR_Resource_Release <PHR> object is "1" and if the value of the "S" field is "0" then the PDR_TTL value included in the <PDR> object must be compared with the calculated PHR_TTL value. When these two values are equal then the local RESERVE (RMD-QSPEC) has to be terminated and it will not be forwarded downstream. The reason of this is that the NF node that is currently processing this message was the last NF node that successfully processed the <PHR> object of its associated initial reservation request (i.e., initial local RESERVE (RMD-QSPEC) message). The next NF downstream node

was unable to successfully process the initial reservation request, and therefore this NF node marked the "M" field of the PHR_Resource_Request <PHR> object. When the values of the "M" and "S" fields are "0" then this message will not be terminated by an NF interior node, but it will be forwarded in the downstream direction. The NF egress node will receive and process the PHR_Resource_Release <PHR> object. Afterwards the NF egress node must terminate the local RESERVE (RMD-QSPEC) object. Future versions of this draft will include more details on this RMD specific release procedure.

5. Bi-directional reservations

Bi-directional reservation is done in the RMD domain by two sender-initiated reservations started from ingress and egress edge.

If QSpec in downlink direction and up-link direction are stacked in the same end-to-end RESERVE message, both QSpec-s are transformed into local QSpec-s in the ingress edge node. The original bi-directional RESERVE is handed over to NF egress. Intra-domain reservation is first done in downlink direction. The uplink local QSpec is encapsulated into the RMD specific downlink RESERVE message. The NF Egress, as a proxy, sends local uplink QSpec to NF ingress in an uplink local reservation message. An uplink reservation message is generated only if positive RESPONSE message arrives at NF egress from the QNR. In order to notify the NF egress about the successful uplink reservation a RESPONSE message has to be sent from NF ingress to NF egress.

If uplik QSPEC is provided by the NR, the NF egress has to translate it to an RMD-QSpec.

In case of unsuccessful reservation, NF ingress and NF egress nodes are notified about the failure, and they initiate "PHR_Release_Request" downlink and uplink directions, respectively, in order to remove unnecessary resources.

Bi-directional reservation will be specified in more detail in a later version of this draft.

6. Fault handling

Fault handling operation refers to the situations when there are problems in the network, such as loss of messages, route change, link failure, etc. Since interior nodes do not store per-flow states the errors are reported to the edge nodes, which make decisions and handle the problem within the domain.

6.1. Message lost

If a reservation or response message is lost, the ingress edge node either refuses connection or re-tries the reservation depending on the policy in the domain after time-out. Since reservation states are soft states, they are also removed after a time-out period.

In case of reservation-based method, refresh or release messages can also be lost. The loss is detected at the edge node that controls the flow times-out and it depends on the network policy how it is handled. One possibility is that the edge node sends again a refresh request message. This solution has the risk of double reservations on certain links. Another possibility is to wait until the next refresh is due to be sent. This solution results that fewer resources will be reserved for almost a full refresh period that may result in QoS violation. A third alternative is to terminate the flow when time-out occurs by explicit release. The basis of this solution is that loss of signaling messages are likely to be caused during severe congestion. Considering the potential loss of release request messages, the soft-state refresh procedure can be used to solve the resulting over-allocation. Future versions of this draft will include more details on this RMD specific procedure.

6.2. Severe congestion

Severe congestion can be considered as an undesirable state, which may occur as a result of a route change but it can be caused by under-estimation of the required resource units. Typically, routing algorithms are able to adapt and change their routing decisions to reflect changes in the topology (e.g., link failures) and traffic volume. In such situations, the re-routed traffic follows a new path. Nodes located on this new path may become overloaded after rerouting. Moreover, when a link fails, the traffic passing through might be dropped, degrading its performance.

Severe congestion occurrence in the communication path has to be notified to the NF edge node that generated the Reservation message. NF Interior node detecting severe congestion marks data packets passing of the node in which the congestion was detected. For severe congestion marking of the data packet, three DSCP-s should be allocated for each traffic class. One is used to indicate that the packet is passed a congested node or not. The other code-point can be used to indicate the degree of congestion. This can be done for example using proportional marking method, which means that the marked bytes are proportional to the degree of congestion. The NF egress node is using a predefined policy to solve the severe congestion, by selecting a number of end-to-end flows that should be terminated. For these flows (sessions), the NF egress node generates and sends a local NOTIFY (PDR) message to the NF ingress node (its previous stateful QoS-NSLP hop) to indicate the severe congestion in the communication path. The SESSION_ID of this message must be the same as the SESSION_ID of the flow that has to be terminated. This message must include a <PDR> object (of type PDR_Reservation_Report). Note that this message will use a NTLP connection mode.

The non-default values of the objects contained in the NOTIFY (PDR) message must be set by the NF egress node as follows:

* the values of the [<ERROR_SPEC>] object is set by the standard QoS-NSLP protocol functions.

* the value of the MsgType field of the PDR object should be set to 8 (PDR_Congestion_Report).

* The value of the "M" field of the <PDR> object must be set to 1.

* The value of the "S" field of the <PDR> object must be set to 1.

* The value of the F-Type depends on the policy used by the NF egress node;

* The value of the EP-Type field of the PDR message should be equal to the QoS-NSLP protocol ID.

* The value of the Flow ID field of the PDR object depends on the policy used by the NF egress node.

Upon receiving this message, the NF ingress node resolves the severe

congestion by a predefined policy, e.g., refusing new incoming flows (sessions), terminating the affected and notified flows (sessions), or shifted to an alternative RMD traffic class (PHB). An example of such an operation is depicted in Figure 7.

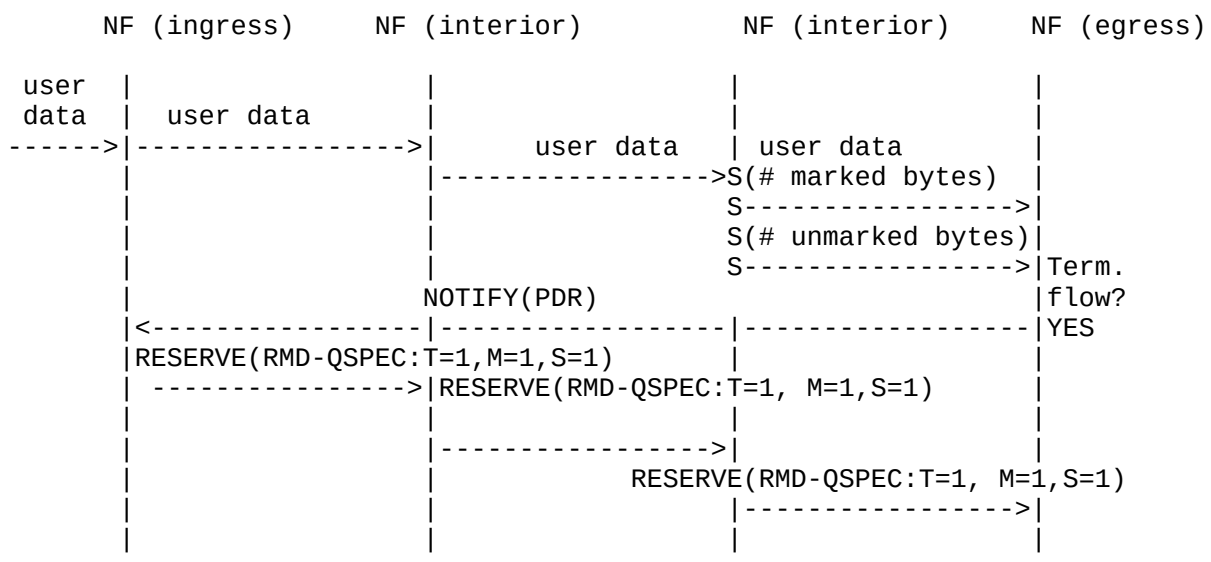


Figure: 7 RMD specific severe congestion handling

Future versions of this draft will include more details on the RMD specific severe congestion procedure.

7. Definition of the RMD QSPEC

Two basic local QoS model objects are defined: Per-hop reservation object (PHR) and per-domain reservation object (PDR). PHR is used for reservation, refresh and release within the domain. PDR is used for edge-to-edge communication, which includes per-domain reservation and response. The QoS model supports both IPv4 and IPv6.

7.1. PHR object

The format of the PHR object for IPv4 and IPv6 versions is depicted in Figure 8.

On top of the PHR specific information of the PHR object, the three (standard) QoS-NSLP object fields are used, i.e., Length, Class-Num and C-type.

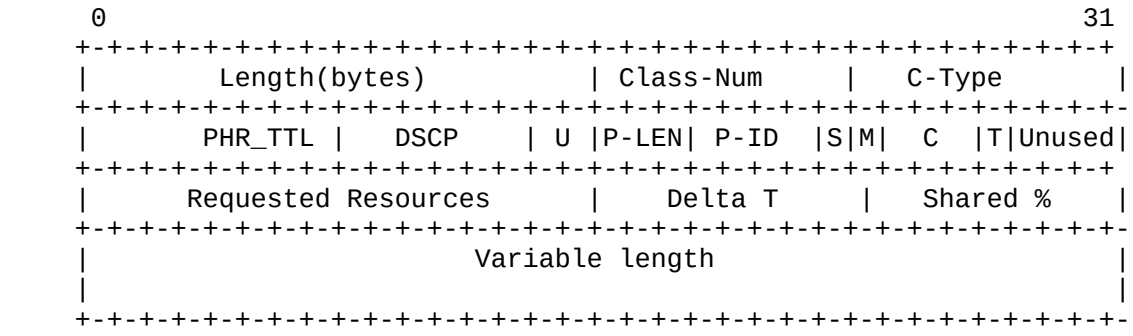


Figure 8: PHR object format

Length (in octets): 16-bit field containing the total object length in octets. It must always be a multiple of 4 and at least 4 octets.

Class-Num: 8-bit field identifying the object class. Each object class has a name. For a QSPEC object this value is for the time being is set to 8.

C-Type: 8-bit field identifying the object type, unique within Class-Num. Object type, unique within Class-Num. For a QSPEC object is specifying the QoS-model ID. For The time being this value is for the RMD-QoS model chosen to be 10.

PHR-TTL: 8-bit field. A counter, that counts the number of RMD reservation based NSIS aware nodes

that could admit (successfully processed) a "PHR_Resource_Request" <PHR> object.

DSCP: 6-bit field. The Diffserv Code Point value used to identify the per traffic class state.

P-LEN (PHR length): 3-bit field. This specifies the length in octets of the specific PHR information data, without including the "Variable length" field. The value 0 specifies that this PHR object contains only data in the "Variable length" field. This data MUST begin on the next 32-bit word boundary after the P-LEN field (after the first "unused" field). In this case, the sender MUST set the "S", "M", "C", and "unused" fields to 0. The P-ID MUST have the value 1. If a receiver receives a packet with a P-LEN value of 0, it MUST ignore the values in the "S", "M", "C", and "unused" fields.

P-ID (PHR type): 4-bit field. This specifies the PHR type. For the reservation based PHR, the value MUST be 1. For the measurement based PHR this value MUST be 2.

S (Severe Congestion): 1-bit field. The sender MUST set the "S" field to 0. This field is set to 1 by an NF(interior) or NF(edge) node when a severe congestion situation occurs.

M (Marked): 1-bit field. The sender MUST set the "M" field to 0. This field is set to 1 by an NF(interior) or NF(edge) node when the node cannot satisfy the "Requested Resources" value.

C (Object type): 3-bit field. This field specifies the type of the PHR object.

C	Description
0	Reserved
1	"PHR_Resource_Request"
2	"PHR_Refresh_Update"
3	"PHR_Release_Request"
4-7	Unused

"PHR_Resource_Request": initiate or update the traffic class reservation state on all nodes located on the communication path between the NF(ingress) and NF(egress) nodes according to an external SAPU Path request.

"PHR_Refresh_Update": refresh the traffic class reservation soft state on all nodes located on the communication path between the NF(ingress) and NF(egress) nodes according to a resource reservation request that was successfully processed by the RSVPv2-NSLP PHR functionality during a previous refresh period.

"PHR_Release_Request": explicitly release, by subtraction, the reserved resources for a particular flow from a traffic class reservation state.

T — 1-bit field. The NF(ingress) node MUST set (TTL de-active) the "T" field to 0. This field MAY be set to "1" by a node when the node will not increase the PHR_TTL value. This is the case when a RMD reservation based NSIS node is not admitting the "PHR_Resource_Request" <PHR> object.

U — A 3-bit field that is currently unused. Reserved for future PHR object extensions.

Delta T — 8 bit field. The value of this field MAY be set by any NF(ingress) node into (only) "PHR_Resource_Release" objects. It specifies a percentage that represents the ratio between a time lag, say T_lag, and the length of the refresh period, say T_period. Where, T_lag represents the difference between the departure time of the previous sent "PHR_Refresh_Update" object and the departure time of the "PHR_Resource_Release" object. T_period represents the length of the refresh period. This information MAY be used by any node during an explicit release procedure.

Shared % — 8 bit field. This value MAY be used to specify if a

(Shared percentage)	load sharing situation occurred on a communication path or not. The ingress node sets this value to 100. If load sharing occurred in a node then the node will divide the shared percentage value to the number of equal cost paths.
Requested Resources	16-bit field. This field specifies the requested number of units of resources to be reserved by a node. The unit is not necessarily a simple bandwidth value. It may be defined in terms of any resource unit (e.g., effective bandwidth) to support statistical multiplexing at message level.
Variable length	this field is currently unused. Reserved for future PHR object extensions.

7.2. PDR object

The format of the PDR object that is based on the IPv4 version is depicted in Figure 9. On top of the PDR specific information of the PDR object, the three (standard) QoS-NSLP object fields are used, i.e., Length, Class-Num and C-type.

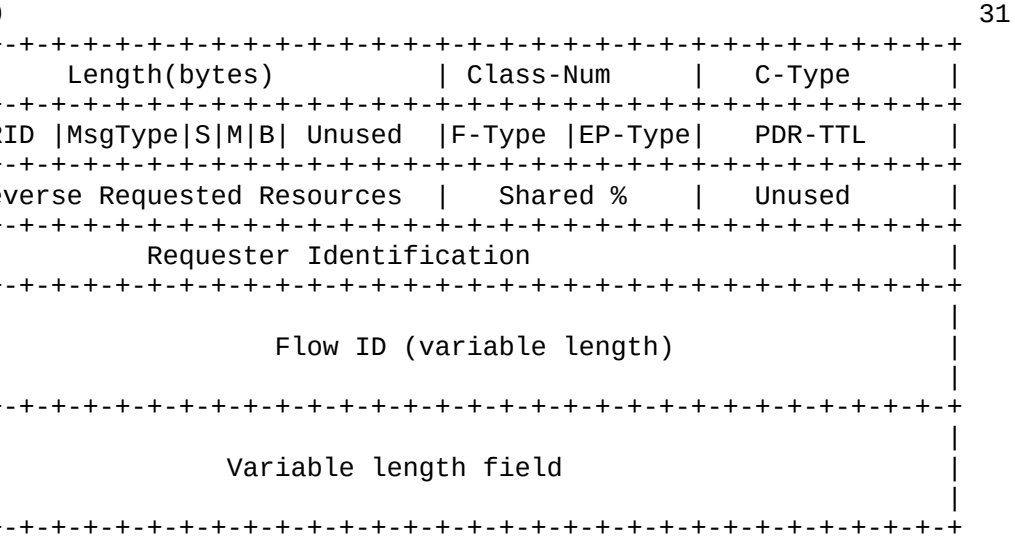


Figure 9: PDR object format for Ipv4

Length (in octets): 16-bit field containing the total object length in octets. It must always be a multiple of 4 and at least 4 octets.

Class-Num: 8-bit field identifying the object class. Each object class has a name. For a QSPEC object this value is for the time being is set to 8.

C-Type: 8-bit field identifying the object type, unique within Class-Num. Object type, unique within Class-Num. For a QSPEC object is specifying the QoS-model ID. For the time being this value is for the RMD-QoS model chosen to be 10.

PDRID: 4-bit field identifying the ID of the PDR object. It is zero for an experimental protocol.

MsgType: 4-bit field identifying the type of PDR object. See below for a table of recognized values.

MsgType	Description	Sent with PHR object
0	reserved	
1	PDR_Reservation_Request	PHR_Resource_Request
2	PDR_Refresh_Request	PHR_Refresh_Update
3	PDR_Release_Request	PHR_Resource_Release
4	PDR_Reservation_Report	
5	PDR_Refresh_Report	
6	PDR_Release_Report	
7	PDR_Request_Info	PHR_Resource_Request OR PHR_Refresh_Update OR PHR_Resource_Release OR PHR_Modification_Request
8	PDR_Congestion_Report	
9	PDR_Modification_Request	
10	PDR_Modification_Report	
11-16	unused	

"PDR_Reservation_Request": generated by the NF(ingress) node in order to initiate or update the QoS-NSLP PDR state in the NF(egress) node

"PDR_Refresh_Request": generated by the NF(ingress) node and sent to the NF(egress) node to refresh, in case needed, the QoS-NSLP PDR states located in the NF(egress) node

"PDR_Modification_Request": generated and sent by the NF(ingress) node to the NF(egress) node to modify the PDR states located in the NF(egress) node

"PDR_Release_Request": generated and sent by the NF(ingress) node to the NF(egress) node to release the flows explicitly

"PDR_Request_Info": an object that can be used

as a common "PDR_Reservation_Request", "PDR_Refresh_Request", "PDR_Release_Request" and "PDR_Modification_Request"

"PDR_Reservation_Report": generated and sent by the NF(egress) node to the NF(ingress) node to report that a "PHR_Resource_Request" PHR object and a "PDR_Reservation_Request" PDR object has been received and that the request has been admitted or rejected

"PDR_Refresh_Report": generated and sent by the NF(egress) node in case needed, to the NF(ingress) node to report that a "PHR_Refresh_Update" PHR object and a "PDR_Refresh_Request" PDR object have been received and have been processed

"PDR_Congestion_Report": generated and sent by the NF(egress) node to the NF(ingress) node and used for Severe congestion notification. They are only used when either the "greedy marking" or "proportional marking" severe congestion notification procedures are applied.

"PDR_Modification_Report": generated and sent by the NF(egress) node to NF(ingress) node to report that the combination of either the "PHR_Resource_Request" PHR object and the "PDR_Modification_Request" PDR object or the "PHR_Release_Request" PHR object and the "PDR_Modification_Request" have been received and processed

PDRID: 4-bit field. ID of the PDR object. It is zero for an experimental protocol.

S (Severe : Congestion) 1-bit field. specifies if a severe congestion situation occured. It can also carry the "S" flag of the "PHR_Resource_Request" or "PHR_Refresh_Update" PHR objects. This flag only applies to "PDR_Reservation_Report", "PDR_Refresh_Report", "PDR_Congestion_Report" and "PDR_Modification_Report" objects.

M (Marked): 1-bit field. Carries the "M" value of the

"PHR_Resource_Request" or "PHR_Refresh_Update" PHR objects.This flag only applies to "PDR_Reservation_Report", "PDR_Refresh_Report", "PDR_Congestion_Report" and "PDR_Modification_Report" objects.

B (Bi-directional reservation) : 1-bit field. specifies that the "PHR" objects should be used for bi-directional reservations in intra-domain signaling. Note that when the inter-domain signaling procedures are applied for bi-directional reservations it does not mean that the associated intra-domain signaling procedures should also use bi-directional reservations.

F-Type: (Flow Type) 4-bit field. The Flow-ID type identifier. Defined by the PDR protocol. It informs the NF(ingress) and NF(egress) nodes what kind of data is contained in the Flow-ID and its length. Every NF(edge) node should be configured to process the F-Types.

EP-Type: (External Protocol Type) 4-bit field. Identifies the used external protocol. If the external protocol is a QoS-NSLP then this field carries the QoS-NSLP protocol ID. Only useful when the intra-domain signaling procedures are used in combination with non-QoS-NSLP inter-domain signaling procedures. It informs the NF(ingress) and NF(egress) nodes what type of external protocol (EP) data is contained in the Variable length field. Every edge node MUST be configured to process the EP-Type. If this field is 0000 then the Variable length field can be used for other purposes, i.e., future specifications.

PDR-TTL: 8-bit field. The PHR_TTL value used to identify the RMD reservation based node that could not admit or process a "PHR_Resource_Request" <PHR> object.

Reverse Requested Resources : 16 bits. This field only applies when the "B" flag is set to "1". It specifies the requested number of units of resources that have to be reserved by a node in the reverse direction when the intra-domain signaling procedures require a

	bi-directional reservation procedure. The unit is not necessarily a simple bandwidth value: it may be defined in terms of any resource unit (e.g., effective bandwidth) to support statistical multiplexing at packet level.
Shared % : (Shared percentage):	8-bit field. This value specifies if a load sharing situation occurred on a communication path or not. The NF(ingress) node sets this value to 100. If load sharing occurred in a node then the node will have to divide the shared percentage value to the number of equal cost paths.
Requester : Identification	32-bit field. For the case that the PDR object is sent by NF(ingress) to NF(egress) this field represents the NF(ingress)Identification. In the other direction this Field represents the NF(egress) identification.
Flow-ID:	Length depends on F-Type. It specifies the flow ID used by the PDR state.
Variable : field	variable length field. It can be used either for length including external protocol data or reserved for future PDR object extensions.

The format of the PDR object that is based on the IPv6 version is depicted in Figure 10. Note that the only difference between the PDR object format based on IPv4 and IPv6 versions is the Requester Identification field, i.e., in IPv6 is this field 128 bits long, while in IPv4 is this field 32 bits long. Note that if RII is used then only the first 32 bits word must be used.

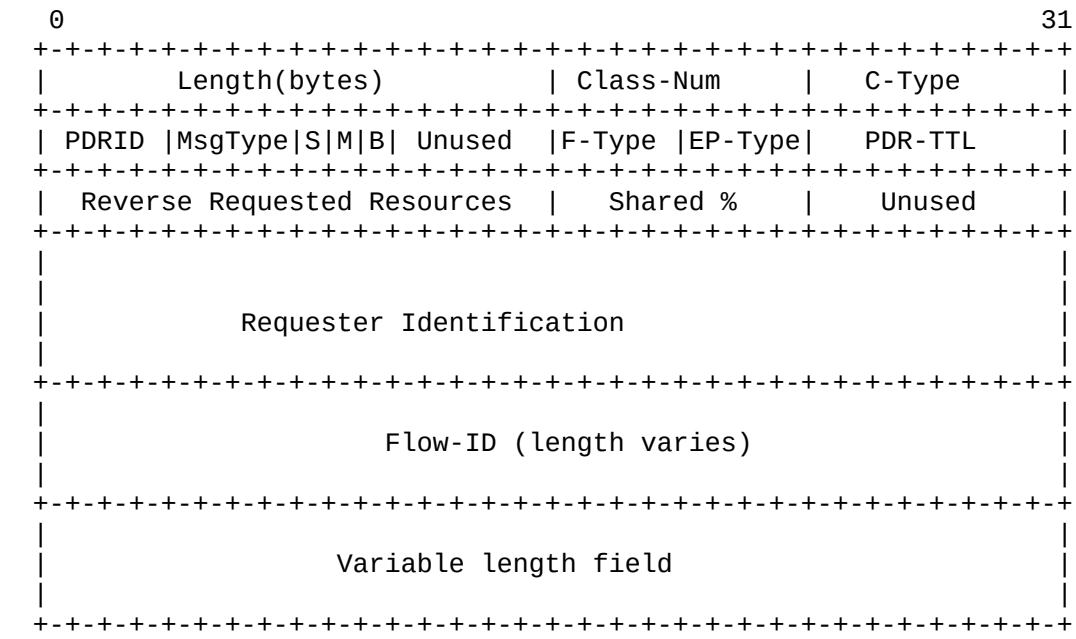


Figure 10: PDR object format based on IPv6

8. Security considerations

Subsequent versions of this draft will need to contain a specification of the RMD QoS model security considerations.

9. Authors' Addresses

Attila Bader
Traffic Lab, Ericsson Research
Ericsson Hungary Ltd.
Laborc u. 1
H-1037 Budapest
Hungary
EMail: Attila.Bader@ ericsson.com

Lars Westberg
Ericsson Research
Torshamnsgatan 23
SE-164 80 Stockholm
Sweden
EMail: Lars.Westberg@ericsson.com

Georgios Karagiannis
University of Twente
P.O. BOX 217
7500 AE Enschede
The Netherlands
EMail: g.karagiannis@ewi.utwente.nl

[Brun03] Brunner, M., "Requirements for Signaling Protocols", IETF Internet Draft, 2003, Work in progress.

[QoS-NSLP] Van den Bosch, S., Karagiannis, G., McDonald, A., "NSLP for Quality-of-Service signalling", IETF Internet Draft, 2003, Work in progress.

[RMD] Westberg, L., et al., "Resource Management in Diffserv (RMD): A Functionality and Performance Behavior Overview", IFIP PfHSN'02, 2002, Berlin. Csaszar, A. et al., "Severe Congestion Handling with Resource Management in Diffserv On Demand", Networking 2002, May 19-24 2002, Pisa - ITALY.; G. Karagiannis, et al., "RMD a lightweight application of NSIS"

[RIMA] Westberg, L., Heijenk, G., Karagiannis, G., Oosthoek, S., Partain, D., Rexhepi, V., Szabo, R., Wallentin, P., el Allali, H.,"Resource Management in Diffserv

Measurement-based Admission Control PHR", Internet draft
Work in progress

[RODA] Westberg, L., Karagiannis, G., Kogel, de M., Partain, D., Oosthoek, S., Jacobsson, M., Rexhepi, V., "Resource Management in Diffserv On DemAnd (RODA) PHR", Internet Draft, Work in progress

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Zh., Weiss, W., "An Architecture for Differentiated Services", IETF RFC 2475, 1998.

[RFC 2961] L. Berger et.al.: "RSVP Refresh Overhead Reduction Extensions", RFC 2961

[RFC 3175] F. Baker C. Iturralde, F. Le Faucheur, B. Davie: "Aggregation of RSVP for IPv4 and IPv6 Reservations", RFC 3175

[GIMPS] Schulzrinne, H., Hancock, R., "GIMPS: General Internet Messaging Protocol for Signaling" Internet Draft, Work in progress.