

NSIS Working Group
INTERNET-DRAFT

Expires: December 2005

Attila Bader
Lars Westberg
Ericsson
Georgios Karagiannis
University of Twente
Cornelia Kappler
Siemens
Tom Phelan
Sonus
June 15, 2005

RMD-QOSM - The Resource Management in Diffserv QoS Model
<[draft-ietf-nsis-rmd-03.txt](#)>

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with [Section 6 of BCP 79](#).

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>.

This Internet-Draft will expire on December 15, 2005.

Copyright Notice

Copyright (C) The Internet Society (2005).

Abstract

This document describes an NSIS QoS Model for networks that use the Resource Management in Diffserv (RMD) concept. RMD is a technique for adding admission control to Differentiated Services (Diffserv) networks. RMD complements the Diffserv architecture by pushing complex classification, conditioning and admission control functions to the edges of a Diffserv domain and simplifying the operation of

internal nodes. The RMD QoS Model allows devices external to the RMD network to signal reservation requests to edge nodes in the RMD network. The RMD Ingress edge nodes classify the incoming flows into

traffic classes and signals resource requests for the corresponding traffic class along the data path to the Egress edge nodes for each flow. Egress nodes reconstitute the original requests and continue forwarding them along the data path towards the final destination.

Table of Contents

1.	Introduction	3
2.	Terminology	3
3.	Overview of RMD and RMD-QOSM	4
3.1	RMD	4
3.2	Basic features of RMD-QOSM	6
3.2.1	Role of the QNEs	6
3.2.2	RMD-QOSM signaling	7
4.	RMD-QOSM, Detailed Description	8
4.1	RMD-QSpec Definition	8
4.1.1	RMD-QOSM QoS Description	9
4.1.2	PHR RMD-QOSM control information	9
4.1.3	PDR RMD-QOSM control information	11
4.1.4	Mapping of QSpec parameters onto generic QSpec Parameters	12
4.2	Message format	13
4.3	RMD node state management	14
4.3.1	Aggregated versus per flow reservations at the QNE edges	14
4.3.2	Measurement-based method	15
4.3.3	Reservation-based method	15
4.4	Transport of RMD-QOSM messages	16
4.5	Edge discovery and addressing of messages	16
4.6	Operation and sequence of events	16
4.6.1	Basic unidirectional operation	17
4.6.1.1	Successful reservation.	17
4.6.1.2	Unsuccessful reservation	21
4.6.1.3	RMD refresh reservation.	23
4.6.1.4	RMD modification of aggregated reservation .	27
4.6.1.5	RMD release procedure.	27
4.6.1.6	Severe congestion handling	33
4.6.2	Bidirectional operation	36
4.6.2.1	Successful and unsuccessful reservation . .	38
4.6.2.2	Refresh reservation	41
4.6.2.3	Modification of aggregated reservation . . .	42
4.6.2.4	Release procedure	43
4.7	Handling of additional errors	46
5.	Security Consideration.	46
6.	IANA Considerations.	48
7.	Open issues.	48
7.1	Explicit congestion notification	48
8.	Acknowledgments.	48

9.	Authors' Addresses.	49
10.	Normative References	50
11.	Informative References	50
12.	Intellectual Property Rights	51

1. Introduction

This document describes a Next Steps In Signaling (NSIS) QoS model for networks that use the Resource Management in Diffserv (RMD) framework ([RMD1], [RMD2], [RMD3]). RMD adds admission control to Diffserv networks and allows nodes external to the networks to dynamically reserve resources within the Diffserv domains.

The Quality of Service NSIS Signaling Layer Protocol (QoS-NSLP) [QoS-NSLP] specifies a generic model for carrying Quality of Service (QoS) signaling information end-to-end in an IP network. Each network along the end-to-end path is expected to implement a specific QoS Model (QOSM) that interprets the requests and installs the necessary mechanisms, in a manner that is appropriate to the technology in use in the network, to ensure the delivery of the requested QoS.

This document specifies an NSIS QoS Model for RMD networks (RMD-QOSM), and an RMD-specific QSpec (RMD-QSPec) for expressing reservations in a suitable form for simple processing by internal nodes. They are used in combination with the QoS-NSLP to provide QoS-NSLP service in an RMD network. Figure 1 shows an RMD network with the respective entities.

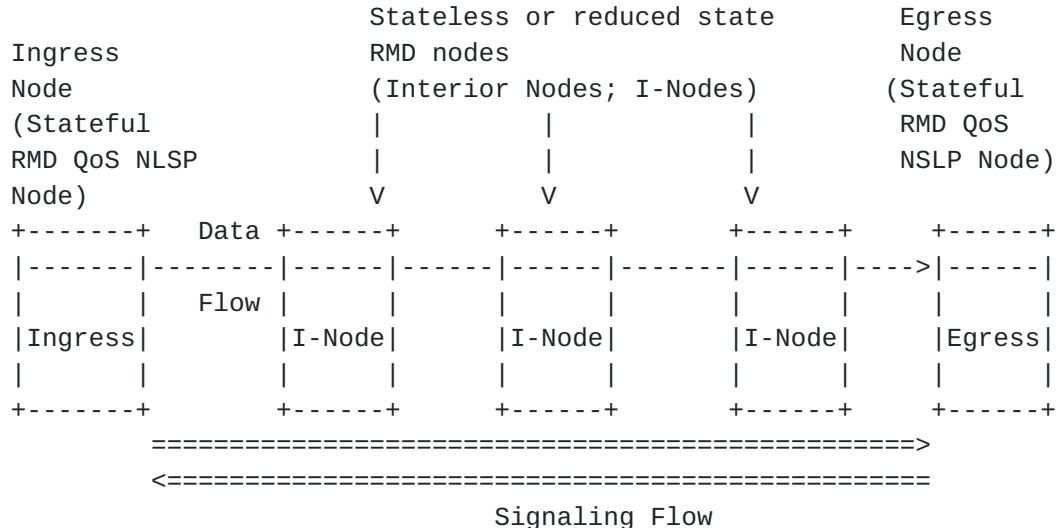


FIGURE 1: Actors in the RMD QOSM

Internally to the RMD network, RMD-QOSM defines a scalable QoS signaling model in which per-flow QoS-NSLP and NTLF states are not stored in Interior nodes but per-flow signaling is performed (see [QoS-NSLP]).

In the RMD-QOSM, only routers at the edges of a Diffserv domain (Ingress and Egress nodes) support the QoS-NSLP stateful operation. Interior nodes support either the QoS-NSLP stateless operation, or a reduced-state operation with coarser granularity than the edge nodes.

The remainder of this draft is structured following the suggestions in [Appendix B](#) of [\[QSP-T\]](#) for the description of QoS Signaling Policies.

After the terminology in [Section 2](#), we give an overview of RMD and the RMD-QOSM in Section 3. In [Section 4](#) we give a detailed description of the RMD-QOSM, including the role of QNEs, the definition of the QSpec, mapping of QSpec generic parameters onto RMD-QOSM parameters, state management in QNEs, and operation and sequence of events. [Section 5](#) discusses security issues.

2. Terminology

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 [RFC 2119](#).

The terminology defined by GIMPS [\[GIMPS\]](#) and QoS-NSLP [\[QoS-NSLP\]](#) applies to this draft.

In addition, the following terms are used:

Edge node: an (NSIS-capable) node on the boundary of some administrative domain.

Ingress node: An edge node that handles the traffic as it enters the domain.

Egress node: An edge node that handles the traffic as it leaves the domain.

Interior nodes: the set of (NSIS-capable) nodes which form an administrative domain, excluding the edge nodes.

3. Overview of RMD and RMD-QOSM

3.1. RMD

The Differentiated Services (Diffserv) architecture ([\[RFC2475\]](#), [\[RFC2638\]](#)) was introduced as a result of efforts to avoid the scalability and complexity problems of Intserv [\[RFC1633\]](#). Scalability is achieved by offering services on an aggregate

rather than per-flow basis and by forcing as much of the per-flow state as possible to the edges of the network. The service

differentiation is achieved using the Differentiated Services (DS) field in the IP header and the Per-Hop Behavior (PHB) as the main building blocks. Packets are handled at each node according to the PHB indicated by the DS field in the message header.

The Diffserv architecture does not specify any way for devices outside the domain to dynamically reserve resources or receive indications of network resource availability. In practice, service providers rely on subscription-time Service Level Agreements (SLAs) that statically define the parameters of the traffic that will be accepted from a customer.

RMD was introduced as a method for dynamic reservation of resources within a Diffserv domain. It describes a method that is able to provide admission control for flows entering the domain and a congestion handling algorithm that is able to terminate flows in case of congestion due to a sudden failure (e.g., link, router) within the domain.

In RMD, scalability is achieved by separating a fine-grained reservation mechanism used in the edge nodes of a Diffserv domain from a much simpler reservation mechanism needed in the Interior nodes. In particular, it is assumed that edge nodes support per-flow QoS states in order to provide QoS guarantees for each flow. Interior nodes use only one aggregated reservation state per traffic class or no states at all. In this way it is possible to handle large numbers of flows in the Interior nodes. Furthermore, due to the limited functionality supported by the Interior nodes, this solution allows fast processing of signaling messages.

In RMD two basic admission control modes are described: measurement-based and reservation-based admission control. The measurement-based algorithm continuously measures traffic levels and the actual available resources, and admits flows whose resource needs are within what is available at the time of the request. Once an admission decision is made, no record of the decision need be kept. The advantage of measurement-based resource management protocols is that they do not require pre-reservation state or explicit release of the reservations. Moreover, when the user traffic is variable, measurement based admission control could provide higher network utilization than, e.g., peak-rate reservation. However, this can introduce an uncertainty in the availability of the resources.

With the reservation-based method, each Interior node maintains only one reservation state per traffic class. The Ingress edge nodes aggregate individual flow requests into classes, and signal changes in the class reservations as necessary. The reservation is quantified in terms of resource units. These resources are

requested dynamically per PHB and reserved on demand in all nodes in the communication path from an Ingress node to an Egress node.

RMD describes the following procedures:

- * classification of individual resource reservation or resource query into Per Hop Behavior groups (PHB) at the Ingress node of the domain,
- * hop-by-hop admission control based on per PHB within the domain. There are two possible modes of operation for internal nodes to admit requests. One mode is the stateless or measurement-based mode, where the resources within the domain are queried. Another mode of operation is the reduced-state reservation or reservation based mode, where the resources within the domain are reserved.
- * a method to forward the original requests across the domain up to the Egress node and beyond.
- * a congestion control algorithm that is able to terminate the appropriate number of flows in case a of congestion due to a sudden failure (e.g., link, router) within the domain.

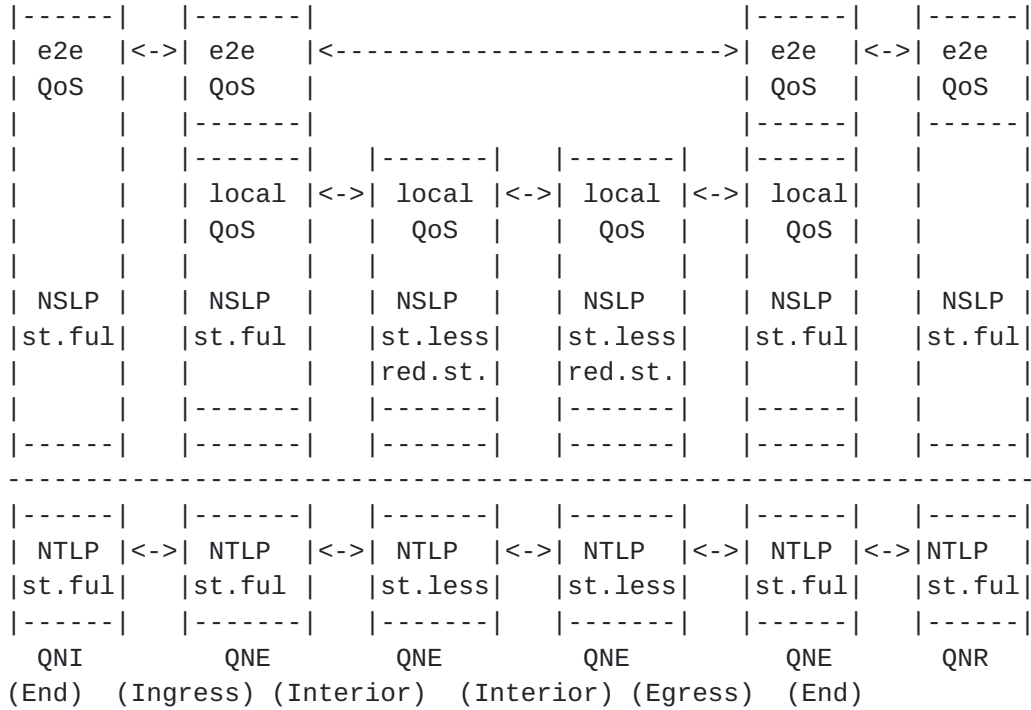
3.2. Basic features of RMD-QOSM

3.2.1 Role of the QNEs

The protocol model of the RMD-QOSM is shown in Figure 2. The figure shows QNI and QNR nodes, not part of the RMD network, that are the ultimate initiator and receiver of the QoS reservation requests. It also shows QNE nodes that are the Ingress and Egress nodes in the RMD domain (QNE Ingress and QNE Egress), and QNE nodes that are Interior nodes (QNE Interior).

All nodes of the RMD domain are QoS-NSLP aware nodes. Edge nodes store and maintain QoS-NSLP and NTLP states and therefore are stateful nodes. The Interior nodes are NTLP stateless. Furthermore they are either QoS-NSLP stateless (for measurement-based operation), or are reduced state nodes storing per PHB aggregated QoS-NSLP states (for reservation-based operation).

Note that the RMD-QOSM domain MAY contain Interior nodes that are not NSIS aware nodes (not shown in the figure). These nodes are assumed to have sufficient capacity for flows that might be admitted. Furthermore, some of these NSIS unaware nodes MAY be used for measuring the traffic congestion level on the data path. These measurements can be used by RMD-QOSM in the severe congestion operation (see [Section 4.6.1.6](#)).



st.ful: stateful, st.less: stateless
st.less red.st.: stateless or reduced state

Figure 2: Protocol model of stateless/reduced state operation

3.2.2 RMD-QOSM signaling

The basic RMD-QOSM signaling is shown in Figure 3. A RESERVE message is created by a QNI with an Initiator QSpec describing the reservation and forwarded along the path towards the QNR. When the original RESERVE message arrives at the Ingress node, an RMD-QSpec is constructed based on the top-most QSPEC in the message (usually the Initiator QSPEC). The RMD-QSpec is sent in a local, independent RESERVE message through the Interior nodes towards the QNR. This local RESERVE message uses the NTLP hop-by-hop datagram signaling mechanism. Meanwhile, the original RESERVE message is sent to the Egress node on the path to the QNR using the reliable transport mode of NTLP.

Each QoS NSLP node on the data path processes the local RESERVE message and checks the availability of resources with either the reservation-based or the measurement-based method. When the message reaches the Egress node, and the reservation is successful in each Interior nodes, the original RESERVE message is forwarded to the next domain. When the Egress node receives a RESPONSE message from the downstream end, it is forwarded directly to the Ingress node.

If an intermediate node cannot accommodate the new request, it indicates this by marking a single bit in the message, and continues forwarding the message until the Egress node is reached. From the Egress node a RESPONSE message is sent directly the Ingress node.

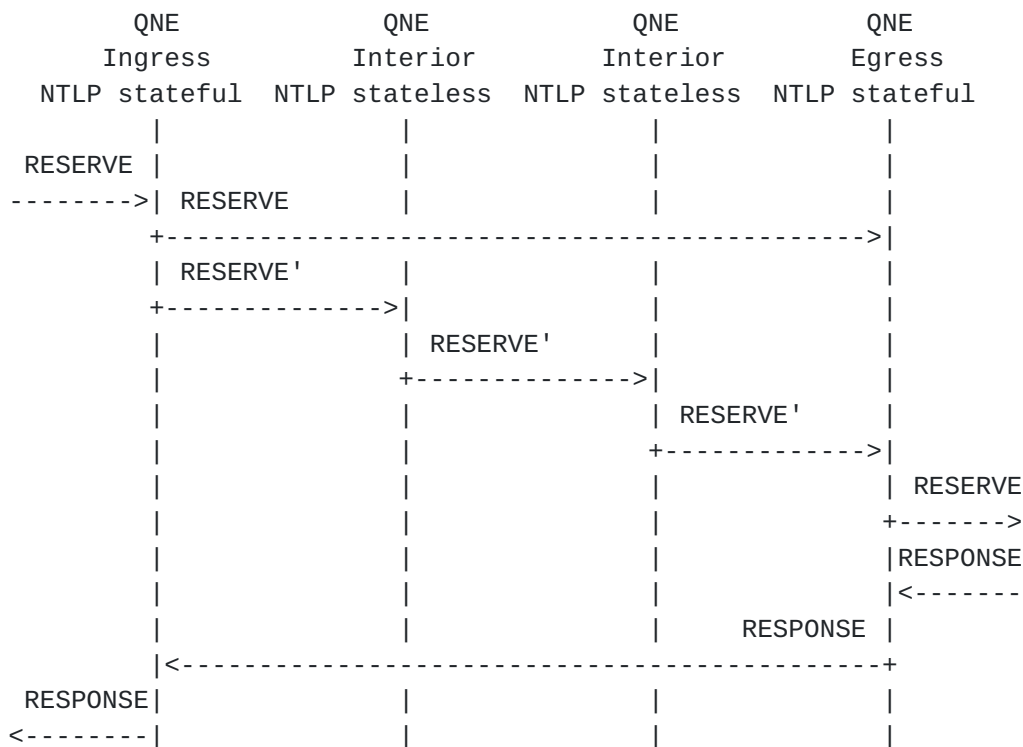


Figure 3: Sender-initiated reservation with Reduced State Interior Nodes

As a consequence in the stateless/reduced state domain only sender-initiated reservation can be performed and functions requiring per flow NTLP or QoS-NSLP states, like summary refreshes, cannot be used. One of the basic features of RMD is that, if per flow identification, is needed, i.e. associating the flows IDs for the reserved resources, Edge nodes act on behalf of Interior nodes.

4. RMD-QOSM, Detailed Description

This section describes RMD-QOSM in more detail. In particular, it defines the role of stateless and reduced-state QNEs, the RMD-QOSM QSpec Object, the format of RMD-QOSM QoS-NSLP messages and how QSpecs are processed and used in different protocol operations.

4.1. RMD-QSpec Definition

The RMD-QOSM QSpec object contains three fields, the "RMD-QOSM QoS Description", the Per Hop Reservation "PHR RMD-QOSM control information" container (PHR container) and the Per Domain Reservation "PDR RMD-QOSM control information" container (PDR container). The "RMD-QOSM QoS Description" field and the "PHR RMD-QOSM control

information" container are used and processed by edge and Interior nodes. The "PDR RMD-QOSM control information" container field is only processed by edge nodes. The "PHR RMD-QOSM control

information" container contains the QoS specific control information for intra-domain communication and reservation. The "PDR RMD-QOSM control information" container contains additional information that is needed for edge-to-edge communication.

4.1.1. RMD-QOSM QoS Description

This section describes the parameters used by the "RMD-QOSM QoS Description" field. The RMD-QOSM QoS Description only contains the QoS Desired object [QSP-T]. It does not contain the QoS Available, QoS Reserved or Minimum QoS objects.

<RMD-QOSM QoS Description> = <QoS Desired>

<QoS Desired> = <Bandwidth> <PHB-CLASS>

The bit format of the <Bandwidth> and <PHB-CLASS> conform to the bit format specified in [QSP-T] and can be seen in Figure 4 and Figure 5, respectively.

```

0                               1                               2                               3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
| Object ID = 1 |Parameter-ID=1 | Length = 5      | Empty          |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|  Bandwidth      (32-bit IEEE floating point number)          |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+

```

Figure 4: Bandwidth parameter

```

0                               1                               2                               3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
| Object ID = 1 |Parameter-ID=9 | Length = 1      | DSCP           |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+

```

Figure 5: PHB_Class parameter

4.1.2. PHR RMD-QOSM control information container (PHR container)

This section describes the parameters used by the PHR container.

<PHR RMD-QOSM control information> = <Overload %>, <S>,<M>,
<Admitted Hops>, , <Hop_U> <Time Lag>

The bit format of the PHR container can be seen in Figure 6. Note that in Figure 6 <Hop U> is represented as <U>.

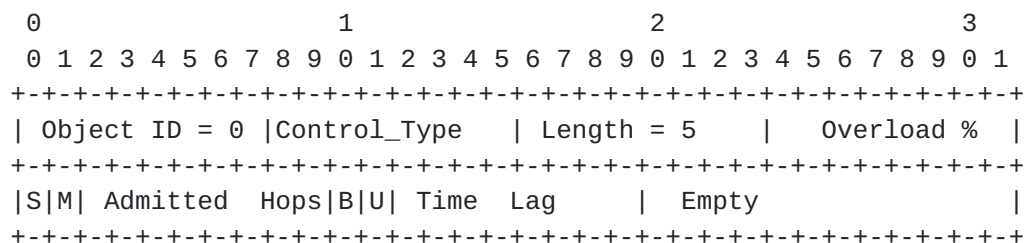


Figure 6: PHR container

Parameter/Container ID:

8 bit field, indicating the PHR type: PHR_Resource_Request, PHR_Release_Request, PHR_Refresh_Update. It is used to further specify QoS-NSLP RESERVE and RESPONSE messages.

"PHR_Resource_Request" (Parameter/Container ID = 1): initiate or update the traffic class reservation state on all nodes located on the communication path between the QNE(Ingress) and QNE(Egress) nodes.

"PHR_Refresh_Update" (Parameter/Container ID = 2): refresh the traffic class reservation soft state on all nodes located on the communication path between the QNE(Ingress) and QNE(Egress) nodes according to a resource reservation request that was successfully processed during a previous refresh period.

"PHR_Release_Request" (Parameter/Container ID = 3): explicitly release, by subtraction, the reserved resources for a particular flow from a traffic class reservation state.

<S> (Severe Congestion):

1 bit. In case of a route change refreshing RESERVE messages follow the new data path, and hence resources are requested there. If the resources are not sufficient to accommodate the new traffic sever congestion occurs. Congested Interior nodes SHOULD notify edge QNEs about the congestion, which is done by setting the S bit.

<0verload %>:

8 bits In case of severe congestion the level of overload is indicated by the Overload %. Overload % SHOULD be higher than 0 if S bit is set. If overload in a node is greater than the overload in a previous node then Overload % SHOULD be updated.

 $\langle M \rangle :$

1 bit. In case of unsuccessful resource reservation or resource query in an Interior QNE, this QNE sets the M bit in order to notify the Egress QNE.

<Admitted Hops>:

8 bit field. The <Admitted Hops> counts the number of hops in the RMD domain where the reservation was successful. The <Admitted Hops> is set to "0" when a RESERVE message enters a domain and increased by one at each Interior QNE. However when a QNE that does not have sufficient resources to admit the reservation is reached, the M Bit is set, and the <Admitted Hops> value is frozen.

<Hop_U> (NSLP_Hops unset):

1-bit. The QNE(Ingress) node MUST set the <Hop_U> parameter to 0. This parameter MAY be set to "1" by a node when the node will not increase the <Admitted Hops> value. This is the case when an RMD-QOSM reservation-based node is not admitting the reservation request. When <Hop_U> is set "1" the <Admitted Hops> SHOULD NOT be changed.

: 1 bit. Indicates bi-directional reservation.

<Time Lag>: 8 bit field. The time lag used in a sliding window over the refresh period.

4.1.3. PDR RMD-QOSM control information container (PDR container)

This section describes the parameters of the PDR container.

The bit format of the PDR container can be seen in Figure 7.

<PDR RMD-QOSM control information> = <Overload %> <S> <M> <Max Admitted Hops> [<PDR Reverse Requested Resources>]

The bit format of the PDR container can be seen in Figure 7. Note that in Figure 7 <Max Admitted Hops> is represented as <Max Adm Hops>.

```

      0              1              2              3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
| Object ID = 0 |Control_Type   | Length = 9   | Overload %   |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|S|M| Max Adm Hops |B| Empty   |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|PDR Reverse Requested Resources(32-bit IEEE floating p.number) |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+

```

Figure 7: PDR container

Parameter/Container ID:

8-bit field identifying the type of PDR container field.

"PDR_Reservation_Request" (Parameter/Container ID = 4): generated by the QNE(Ingress) node in order to initiate or update the QoS-NSLP per domain reservation state in the QNE(Egress) node

"PDR_Refresh_Request" (Parameter/Container ID = 5): generated by the QNE(Ingress) node and sent to the QNE(Egress) node to refresh, in case needed, the QoS-NSLP per domain reservation states located in the QNE(Egress) node

"PDR_Release_Request" (Parameter/Container ID = 6): generated and sent by the QNE(Ingress) node to the QNE(Egress) node to release the per domain reservation states explicitly

"PDR_Reservation_Report" (Parameter/Container ID = 7): generated and sent by the QNE(Egress) node to the QNE(Ingress) node to report that a "PHR_Resource_Request" and a "PDR_Reservation_Request" control information fields have been received and that the request has been admitted or rejected

"PDR_Refresh_Report" (Parameter/Container ID = 8) generated and sent by the QNE(Egress) node in case needed, to the QNE(Ingress) node to report that a "PHR_Refresh_Update" control information field has been received and has been processed

"PDR_Release_Report" (Parameter/Container ID = 9) generated and sent by the QNE(Egress) node in case needed, to the QNE(Ingress) node to report that a "PHR_Release_Request" and a "PDR_Release_Request" control information fields have been received and have been processed.

"PDR_Congestion_Report" (Parameter/Container ID = 10): generated and sent by the QNE(Egress) node to the QNE(Ingress) node and used for Severe congestion notification

<S> (PDR Severe Congestion):

1-bit. Specifies if a severe congestion situation occurred. It can also carry the <S> parameter of the "PHR_Resource_Request" or "PHR_Refresh_Update" fields.

<Overload %>:

8-bit. It includes the Overload % of the "PHR_Resource_Request" or "PHR_Refresh_Update" control information fields, indicating the level of overload to the Ingress node.

<M> (PDR Marked):

1-bit. Carries the <M> value of the "PHR_Resource_Request" or "PHR_Refresh_Update" control information fields.

: 1 bit Indicates bi-directional reservation.

<Max Admitted Hops>:

8-bit. The <Admitted Hops> value that has been carried by the PHR container field used to identify the RMD reservation based node that admitted or process a "PHR_Resource_Request"

<PDR Reverse Requested Resources>

32 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.

4.2. Message format

The format of the messages used by the RMD-QOSM 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.

The format of a local (intra-domain) RESERVE message used by the RMD-QOSM is:

```
RESERVE = COMMON_HEADER
          RSN [ RII ] [ REFRESH_PERIOD ] [ BOUND_SESSION_ID ]
          [ POLICY_DATA ] [ RMD-QSPEC ]
```

The format of a Query message used by the RMD-QOSM is as follows:

```
QUERY = COMMON_HEADER
        [ RII ][ BOUND_SESSION_ID ]
        [ POLICY_DATA ] [ RMD-QSPEC ]
```

A QUERY message MUST contain an RII object to indicate a RESPONSE is desired, unless the QUERY is being used to initiate reverse-path state for a receiver-initiated reservation.

The format of a local (intra-domain) RESPONSE message used by the RMD-QOSM is as follows:

```
intra-domain RESPONSE = COMMON_HEADER
                        [ RII / RSN ] ERROR_SPEC
```


The format of an end-to-end RESPONSE message that is used by the RMD-QOSM to carry an intra-domain RESPONSE message is as follows:

```
RESPONSE = COMMON_HEADER  
           [ RII / RSN ] ERROR_SPEC [ RMD-QSPEC ] [ *QSPEC ]
```

The format of a NOTIFY message used by the RMD-QOSM is as follows:

```
NOTIFY = COMMON_HEADER ERROR_SPEC [ RMD-QSPEC ]
```

All objects, except RMD-QSPEC objects, are specified in [\[QoS-NSLP\]](#).

[4.3.](#) RMD node state management

The QoS-NSLP state creation and management is specified in [\[QoS-NSLP\]](#). This section describes the state creation and management functions of the Resource Management Function (RMF) in the RMD nodes.

[4.3.1](#) Aggregated versus per flow reservations at the QNE edges

The QNE edges maintain for the RMD QoS model either per flow, or aggregated QoS-NSLP reservation states. Each per flow or aggregated QoS-NSLP reservation state, associated with the RMD-QoS model, is identified by a NTLF SESSION_ID (see [\[GIMPS\]](#)). In RMD, these states are denoted as PDR states.

In the situation where the QNE edges maintain per aggregated QoS-NSLP reservation states then these states will have to maintain the SESSION_ID of the aggregated state, the IP addresses of the Ingress and Egress nodes, the PHB value and the size of the aggregated reservation, e.g., reserved bandwidth.

The size of the aggregation is defined as it is specified in [Section 1.4.4 of \[RFC 3175\]](#). The size of the aggregated reservations needs to be greater or equal to the sum of bandwidth of the inter domain (end-to-end) reservations it aggregates. Some policy can be used to maintain the amount of required bandwidth on a given aggregated reservation by taking into account the sum of the underlying inter domain (end-to-end) reservations, while endeavoring to change reservation less frequently. This MAY require a trend analysis. If there is a significant probability that in the next interval of time the current aggregated reservation is exhausted, the Ingress router MUST predict the necessary bandwidth and request it. If the Ingress router has a significant amount of bandwidth reserved but has very little probability of using it, the policy MAY predict the amount of bandwidth required and release the excess. To increase or

decrease the aggregate, the RMD modification procedures SHOULD be used (see [Section 4.6.1.4](#)).

4.3.2 Measurement-based method

QNE Interior nodes operating in measurement-based mode are QoS-NSLP stateless nodes, i.e., they do not support any QoS-NSLP or NTLP/GIMPS states. These measurement-based nodes do store two RMD-QOSM states per PHR group. These states reflect traffic conditions at the node and are not affected by any QoS-NSLP signaling. One state stores the measured user traffic load associated with the PHR group and another state stores the maximum traffic load that can be admitted per PHR group.

When a measurement-based node receives a local RESERVE message, it compares the requested resources to the available resources (maximum allowed minus current load) for the requested PHR group. If there are insufficient resources, it sets the <M> bit in the RMD-QSpec. No change to the RMD-QSpec is made when there are sufficient resources. In either case, the node then forwards the RESERVE along the path towards the destination. REFRESH and RELEASE messages are not normally generated in the measurement-based mode, but if received SHOULD not be processed and forwarded unchanged.

4.3.3 Reservation-based method

QNE Interior nodes operating in reservation-based mode are QoS-NSLP reduced state nodes, i.e., they do not store NTLP/GIMPS states but they do store per-PHB-aggregated QoS-NSLP states.

The reservation-based PHR installs and maintains one reservation state per PHB, in all the nodes located in the communication path from the QNE Ingress node up to the QNE Egress node. This state represents the number of currently reserved resource units. Thus, the QNE Ingress node signals only the resource units requested by each flow. These resource units if admitted are 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 are soft states, which are refreshed by sending periodic refresh local RESERVE messages. If a refresh message corresponding to a number of reserved resource units is not received, the aggregated reservation state is decreased in the next refresh period by the corresponding amount of resources that were not refreshed. The refresh period can be refined using a sliding window algorithm described in [RMD3].

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.

4.4. Transport of RMD-QOSM messages

The intra-domain (local) messages used by the RMD-QOSM MUST operate in the NTLP/GIMPS Datagram mode (see [\[GIMPS\]](#)). Therefore, the NSLP functionality available in all QoS NSLP nodes that are able to support the RMD-QOSM MUST require the intra-domain GIMPS functionality available in these nodes to operate in the datagram mode, i.e., require GIMPS to:

- * operate in unreliable mode. This can be satisfied by passing this requirement from the QoS-NSLP layer to the GIMPS layer via the API transfer-attributes.
- * do not create a message association state. This requirement can be satisfied by a local policy, e.g., the QNE is configured to do not create a message association state
- * do not create any NTLP routing state. This can be satisfied by passing this requirement from the QoS-NSLP layer to the GIMPS layer via the API.

All the intra-domain local messages are transported using the GIMPS data messages (see [\[GIMPS\]](#)).

4.5 Edge discovery and addressing of messages

Mainly, the Egress node discovery can be performed either by using the GIMPS discovery mechanism [\[GIMPS\]](#), manual configuration or any other discovery technique. The addressing of signaling messages depends on the used GIMPS transport mode. The RMD QoS signaling messages that are processed only by the edge nodes use the peer-peer addressing of the GIMPS connection mode (C). RMD QoS signaling messages that are processed by all nodes of the Diffserv domain, i.e., edges and Interior nodes, use the end-end addressing of the GIMPS datagram (D) mode. RMD QoS signaling messages addressed to the the data path end nodes are intercepted by the Egress nodes.

4.6. Operation and sequence of events

This section describes the operation and the sequence of events in

the RMD-QOSM.

Bader, et al.

[Page 16]

4.6.1. Basic unidirectional operation

This section describes the basic unidirectional operation and sequence of events of the RMD-QOSM. The following basic operation cases are distinguished: Successful reservation, Unsuccessful reservation, Refresh, Modification, Release and Severe congestion. The QNEs at the edges of the RMD domain support the RMD QoS Model and end-to-end QoS models, which process the RESERVE message differently. Note that the term end-to-end QoS model applies to any QoS model that is initiated and terminated outside the RMD-QOSM aware domain. However, there might be situations where a QoS model is initiated and/or terminated by the QNE edges and is considered to be an end-to-end QoS model. This can occur when the QNE edge can also operate as a QNI or as a QNR.

4.6.1.1. Successful reservation

This section describes the operation of the RMD-QOSM where a reservation is successfully accomplished.

The QNI generates the initial RESERVE message, and it is forwarded by the NTLP as usual [[GIMPS](#)].

4.6.1.1.1. Operation in Ingress node

When an end-to-end reservation request (RESERVE) arrives at the Ingress node (QNE), see Figure 8, it is processed based on the procedures defined by the end-to-end QoS model. Subsequently, the RMD QoS Description: <Bandwidth> and <PHB-CLASS> are derived from the QoS Description of the end-to-end QSpec.

As described in [Section 4.3.1](#), the QNE edges maintain for the RMD QoS model either per flow, or aggregated QoS-NSLP reservation states, which are identified by (local NTLP) SESSION_IDs (see [[GIMPS](#)]). Note that this NTLP SESSION ID is a different one than the SESSION_ID associated with the end-to-end RESERVE message.

If the request was satisfied locally (see [Section 4.3](#)), the Ingress QNE node generates two RESERVE messages: one intra-domain and one end-to-end RESERVE messages. These are bounded together including BOUND_SESSION_ID in the intra-domain RESERVE message.

The intra-domain RESERVE message is associated with the (local NTLP) SESSION ID mentioned above. The selection of the IP source and IP destination address of this message depends on if and how the different inter domain (end-to-end) flows can be aggregated by the QNE Ingress node (see [Section 4.3.1](#)). If no QOS-NSLP aggregation procedure at the QNE edges is possible then the IP source and IP destination address of this message MUST be equal to the IP Source

and IP destination addresses of the data flow. The intra-domain
RESERVE message must be sent using the NTLP datagram mode, see

[Section 4.4](#). In addition, the intra-domain RESERVE (RMD-QSPEC) message MUST include a PHR container (PHR_Resource_Request) and the "RMD QoS Description" field.

The end-to-end RESERVE message includes the end-to-end QSpec and it is sent to the Egress QNE. If the end-to-end QSpec does not carry an RII object, then the A (Acknowledgment) flag MUST be set ON. Otherwise the A flag MUST be set OFF.

Note that after completing the initial discovery phase, the GIMPS connection mode between the QNE Ingress and QNE Egress can be used. The end-to-end RESERVE message is forwarded using the GIMPS forwarding procedure to bypass the Interior stateless or reduced-state QNE nodes, see Figure 8. Furthermore, note that the initial discovery phase and the process of sending the end-to-end RESERVE message towards the QNE Egress MAY be accomplished simultaneously.

The (initiating) intra-domain RESERVE message MUST be used and/or set by the QNE Ingress 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 value of the <BOUND_SESSION_ID> object MUST be the session ID associated to the end-to-end RESERVE message;
- * the SCOPING flag SHOULD not be set, meaning that a default scoping of the message is used. Therefore, the QNE edges MUST be configured as boundary nodes and the QNE Interior nodes MUST be configured as Interior (intermediary) nodes;
- * The <RII> object is not included in this message;
- * the value of the <REFRESH_PERIOD> object MUST be calculated and set by the QNE Ingress node, see also [Section 4.6.1.3](#);
- * the PHR resource units MUST be included into the <Bandwidth> parameter of the "RMD QoS Description" field;
- * the value of the Parameter/Container ID field of the PHR container MUST be set to 1, (i.e., PHR_Resource_Request;)
- * the value of the <Admitted Hops> parameter in the PHR container MUST be set to "1";
- * the value of the <Hop_U> parameter in the PHR container MUST be set to "0";
- * the flag "Acknowledge" (A) MUST be set "OFF";

- * In a single-domain case the PDR container MAY not be included into the message.

When an end-to-end RESPONSE(PDR) message is received by the QNE Ingress node, the RMD-QSPEC, see [Section 4.6.1.1.3](#), has to be identified, processed and removed from the end-to-end RESPONSE message. The QoS-NSLP state in the QNE Ingress stores and maintains the binding between each end-to-end session and each intra-domain session. In this way the QNE Ingress can match the PHR container that has been carried by the intra-domain RESERVE with the received PDR container that has been carried by the end-to-end RESPONSE message. The RMD QoS model functionality is notified by reading the <M> parameter of the "PDR RMD control information" container that the reservation has been successful.

If the end-to-end RESPONSE message has to be forwarded to a node outside the RMD-QOSM aware domain then the non-default values of the objects contained in this message (i.e., <RII/RSN>, <ERROR_SPEC>, [*QSPEC]) MUST be used and set by the QoS-NSLP protocol functions of the QNE.

4.6.1.1.2 Operation in the Interior nodes

Each QNE Interior node MUST use the QoS-NSLP and RMD-QOSM parameters of the intra-domain RESERVE (RMD-QSPEC) message as follows:

- * the values of the <RSN>, <RII>, <REFRESH_PERIOD>, <BOUND_SESSION_ID>, <POLICY_DATA> objects are not changed, i.e., equal to the values set by the QNE Ingress. These values are not used by the QNE Interior;
- * the flag "Acknowledge" (A) SHOULD be set "OFF";
- * the value of <Bandwidth> parameter of the "RMD QoS Description" field is used by the QNE Interior node for admission control, see [Section 4.3.2](#) and [Section 4.3.3](#);
- * in case of the RMD reservation-based procedure, and 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 <Admitted Hops> parameter in the PHR container has to be increased by one;
- * in case of the RMD measurement based method, and if these resources are admitted, using a MBAC algorithm, the number of this resources will be used to update the MBAC algorithm.

4.6.1.1.3 Operation in the Egress node

When the intra-domain RESERVE(RMD-QSPEC) is received by the QNE Egress node of the session associated with the intra-domain RESERVE(RMD-QSPEC) (the PHB session) with the session included in its <BOUND_SESSION_ID> object MUST be bounded. The session included in the <BOUND_SESSION_ID> object is the session associated with the end-to-end RESERVE message.

The end-to-end RESERVE message is only forwarded further, towards QNR, if the processing of the intra-domain RESERVE (RMD-QSPEC) message was successful at all nodes in the RMD domain. Otherwise the inter domain (end-to-end) reservation is considered as being failed.

If the (A) flag carried by the end-to-end RESERVE message was set to ON, then a one hop (end-to-end) RESPONSE message MUST be generated by the QNE Egress. Otherwise, the QNE Egress MUST wait for the end-to-end RESPONSE message that has the same SESSION ID as the end-to-end RESERVE message forwarded towards QNR.

The non-default values of the objects contained in the end-to-end RESPONSE(PDR) message MUST be used and/or set by the QNE Egress as follows:

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

In addition to the above, the QNE Egress MUST also generate a RMD-QSPEC object that is carried by the end-to-end RESPONSE (PDR) message, see [Section 4.2](#).

The following parameters of the RMD-QSPEC object MUST be used and/or set in the following way:

- * the value of the Parameter/Container ID field of the PDR container MUST be set "7" (i.e., PDR_Reservation_Report);
- * the value of the <M> field of the PDR container MUST be equal to the value of the <M> parameter of the PHR container that was carried by its associated intra-domain RESERVE (RMD-QSPEC) message.

The end-to-end RESPONSE (PDR) message is addressed and sent to its upstream QoS-NSLP neighbor, i.e., QNE Ingress node. Note that for all upstream messages the RAO is not set. Therefore, all Interior nodes ignore the end-to-end RESPONSE messages.

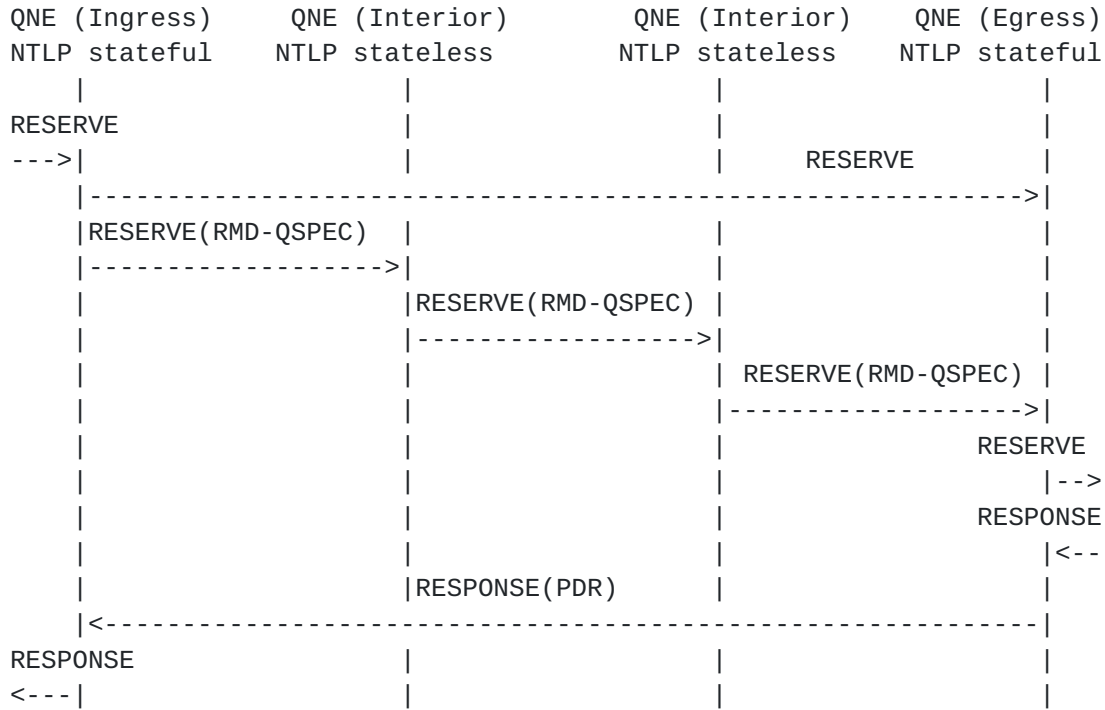


Figure 8: Basic operation of successful reservation procedure used by the RMD-QOSM

4.6.1.2. Unsuccessful reservation

This section describes the operation where a request for reservation cannot be satisfied by the RMD-QOSM.

The QNE Ingress, the QNE Interior and QNE Egress nodes process and forward the end-to-end RESERVE message and the intra-domain RESERVE (RMD-QSPEC) message in the same way as specified in [Section 4.6.1.1](#). The main difference between the unsuccessful operation and successful operation is that one of the QNE nodes does not admit the request due to lack of resources. This also means that the QNE edge node MUST NOT forward the end-to-end RESERVE message towards the QNR node.

4.6.1.2.1 Operation in the Ingress nodes

When an end-to-end RESERVE message arrives to the QNE Ingress and if there are no resources available locally, the QNE Ingress MUST reject this end-to-end RESERVE message and sends a RESPONSE message back to the sender, using a standard QoS-NSLP procedure.

In case of the RMD reservation based scenario, and if the intra-domain reservation request is not admitted by the QNE Interior node then the <Hop_U> and <M> parameters of the PHR container MUST be

set to "1". The <Admitted Hops> counter MUST NOT be increased.

Bader, et al.

[Page 21]

In case of the RMD measurement based scenario, and if the intra-domain reservation query (i.e., intra-domain RESERVE(RMD-QSPEC) is not admitted by the MBAC algorithm then the <M> parameter of the PHR container MUST be set to "1".

When an end-to-end RESPONSE(PDR) message is received by an Ingress node, see [Section 4.6.1.2.3](#), the following actions take place. The non-default values of the objects contained in the end-to-end RESPONSE (PDR) message MUST be used and/or set by the QNE Ingress node as follows:

- * the values of the <RII/RSN>, <ERROR_SPEC>], [*QSPEC] objects are set by standard QoS-NSLP protocol functions
- * the RMD-QSPEC object, see [Section 4.2](#), has to be processed and removed. The RMD Resource Management Function (RMF) is notified by reading the <M> parameter of the PDR container that the reservation has been unsuccessful. In case of a RMD reservation based scenario, the RMD-QOSM functionality, has to start an RMD release procedure (see [Section 4.6.1.5](#)).

[4.6.1.2.2](#) Operation in the Interior nodes

In general, if a QNE Interior node receives a PHR container, of type "PHR_Resource_Request", with the <M> parameter set to "1" then this PHR container and the "RMD QoS Description" field MUST NOT be Processed.

[4.6.1.2.3](#) Operation in the Egress nodes

In the RMD reservation based and RMD measurement based scenario, when the <M> marked intra-domain RESERVE(RMD-QSPEC) is received by the QNE Egress node (see Figure 9) the session associated with the intra-domain RESERVE(RMD-QSPEC) (the PHB session) and the session included in its BOUND_SESSION_ID object MUST be bounded. The session included in the <BOUND_SESSION_ID> object is the session associated with the end-to-end RESERVE.

The QNE Egress node MUST generate an end-to-end RESPONSE message that will have to be sent to its previous stateful QoS-NSLP hop.

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

In addition to the above, and similar to the successful operation, see [Section 4.6.1.1.3](#), the QNE Egress MUST also generate an RMD-QSPEC object that is carried by the end-to-end RESPONSE message.

The following fields of the RMD-QSPEC object MUST be used and/or set in the following way:

- * the value of the <PDR Control Type> of the PDR container MUST be set to "7" (PDR_Reservation_Report);
- * the value of the <Admitted Hops> parameter of the PHR container included in the received <M> marked PDR container MUST be included in the <Max_Admitted Hops> parameter of the PDR container;
- * the value of the <M> parameter of the PDR container MUST be set to "1".

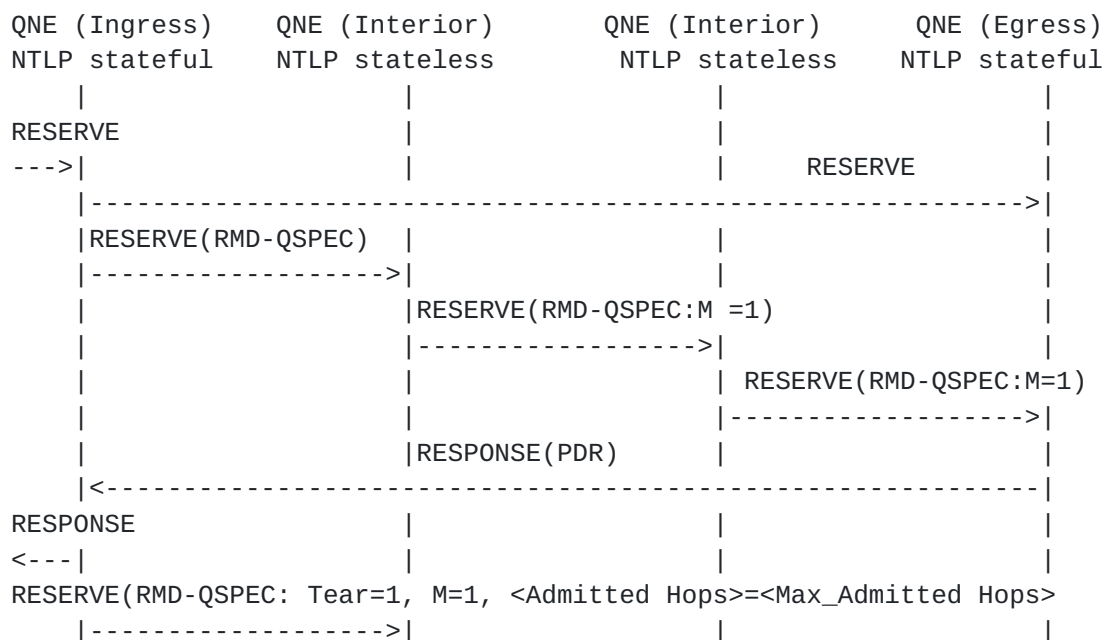


Figure 9: Basic operation during unsuccessful reservation initiation used by the RMD-QOSM

4.6.1.3 RMD 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 the QNE Egress.

The refresh procedure in case of RMD reservation-based method follows a similar scheme as the reservation process, shown in Figure 3. If the RESERVE messages arrive within the soft state time-out period, the corresponding number of resource units are not removed. However, the transmission of the intra-domain and end-to-end (refresh) RESERVE message are not necessarily synchronized. Furthermore, the generation of the end-to-end RESERVE message, by the QNE edges, depends on the locally maintained refreshed interval (see [[QoS-NSLP](#)]).

4.6.1.3.1 Operation in the Ingress node

The Ingress node MUST be able to generate an intra-domain (refresh) RESERVE (RMD-QSpec) at any time. Before generating this message, the RMD QoS signaling 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 MUST be equal in all QNE edge and QNE Interior nodes and SHOULD be smaller (default: more than two times) than the refresh period at the QNE Ingress node used by the end-to-end RESERVE message.

The intra-domain RESERVE (RMD-QSPEC) message MUST include a "RMD QoS Description" field and a PHR container (i.e. PHR_Refresh_Update).

The selection of the IP source and destination address of this message depends on if and how the different inter domain (end-to-end) flows can be aggregated by the QNE Ingress node (see [Section 4.3.1](#)). Note that this QoS-NSLP aggregation procedure is different than the RMD traffic class aggregation procedure. One example is the approach used by the RSVP aggregation scenario ([[RFC 3175](#)]), where the IP source address of this message is the IP address of the aggregator (i.e., QNE Ingress) and the IP destination address of this message is the IP address of the De-aggregator (i.e., QNE Egress). Another example approach is the one used in "RSVP Refresh Overhead Reduction Extensions" ([[RFC2961](#)]). If no QoS-NSLP aggregation procedure at the QNE edges is possible then the IP source and IP destination address of this message MUST be equal to the IP source and IP destination addresses of the data flow.

An example of this RMD specific refresh operation can be seen in Figure 10.

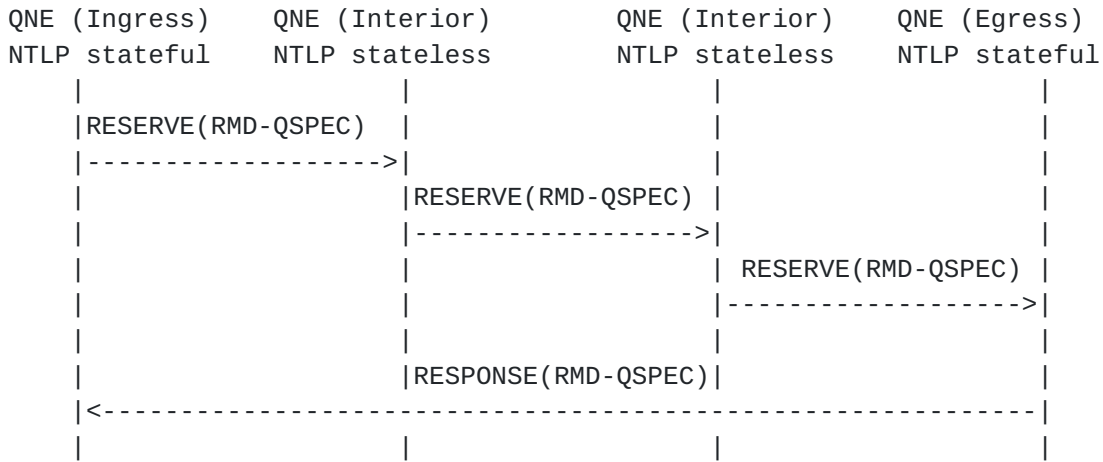


Figure 10: Basic operation of RMD specific refresh procedure

Most of the non-default values of the objects contained in this message MUST be used and/or set by the QNE Ingress in the same way as described in [Section 4.6.1.1](#). The following objects are used and/or set differently:

- * the flag "Acknowledge" (A) SHOULD be set "OFF"
- * the PHR resource units MUST be included into the <Bandwidth> parameter. The value of the <Bandwidth> parameter depends on how the different inter domain (end-to-end) flows are aggregated by the QNE Ingress node (e.g., the sum of all the PHR requested resources of the aggregated flows). If no QoS-NSLP aggregation is accomplished by the QNE Ingress node, the value of the <Bandwidth> parameter SHOULD be equal to the <Bandwidth> parameter of its associated new (initial) intra-domain RESERVE (RMD-QSPEC) message;
- * the value of the Parameter/Container field of the "PHR RMD-QOSM control information" container MUST be set to "2", i.e., "PHR_Refresh_Update";
- * In a single-domain case the PDR container field MAY not be included into the message.
- * the value of the <RII> object MUST contain the Response Identification Information value of the Ingress QNE, that is unique within a session and different for each message (see [\[QoS-NSLP\]](#)).

When the intra-domain RESPONSE (RMD-QSPEC) message, see [Section 4.6.1.3.3](#), is received by the QNE Ingress node, then:

- * the values of the <RII/RSN>, <ERROR_SPEC>, [*QSPEC] objects

are processed by the standard QoS-NSLP protocol functions;

- * the PDR has to be processed and removed by the RMD-QOSM functionality in the QNE Ingress node. The RMD-QOSM functionality is notified by the <PDR M> parameter of the PDR container 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 QNE Ingress node has to generate (in a standard QoS-NSLP way) an error end-to-end RESPONSE message that will be sent towards QNI.

4.6.1.3.2 Operation in the Interior node

The intra-domain RESERVE (RMD-QSPEC) message is received and processed by the QNE Interior nodes. Any QNE edge or QNE Interior node that receives a "PHR_Refresh_Update" control information field MUST identify the traffic class state (PHB) (using the <PHB-CLASS> parameter). Most of the parameters in this refresh intra-domain RESERVE (RMD-QSPEC) message MUST be used and/or set by a QNE Interior node in the same way as described in [Section 4.6.1.1](#).

The following objects are used and/or set differently:

- * the value of <Bandwidth> parameter of the "RMD QoS Description" field is used by the QNE Interior node for refreshing the RMD traffic class state. These resources (included in <Bandwidth>), if reserved, are added to the currently reserved resources per PHB and therefore they will become a part of the per traffic class (per-PHB) reservation state. If the refresh procedure cannot be fulfilled then the <M> parameter of the PHR container has to be set to "1".

Any PHR container of type "PHR_Refresh_Update", and its associated "RMD QoS Description" field (i.e., <Bandwidth>), whether it is marked or not, is always processed, but marked bits are not changed.

4.6.1.3.3 Operation in the Egress node

The intra-domain RESERVE(RMD-QSPEC) message is received and processed by the QNE Egress node. A new intra-domain RESPONSE (RMD-QSPEC) message is generated by the QNE Egress node. This message MUST include a PDR (type PDR_Refresh_Report).

The intra-domain RESPONSE (RMD-QSPEC) message MUST be sent to the QNE Ingress node, i.e., previous stateful hop. The address of the QNE Ingress node can be found using the existing messaging association between the QNE Egress and QNE Ingress nodes. This state is associated with the end-to-end session and identified by the SESSION ID that is bound to the session of the intra-domain RESPONSE(RMD-QSPEC) message.

The following objects MUST be used and/or set differently:

- * the value of the <PDR Control Type> parameter of the PDR container MUST be set "8" (i.e. PDR_Refresh_Report).

4.6.1.4. RMD modification of aggregated reservations

In the case when the QNE edges maintain QoS-NSLP aggregated reservation states and the aggregated reservation has to be modified (see [Section 4.3.1](#)) the following procedure is applied:

- * When the modification request requires an increase of the reserved resources, the QNE Ingress node MUST include the corresponding value into the <Bandwidth> parameter of the "RMD QoS Description" field, which is sent together with a "PHR_Resource_Request" control information. If a QNE edge or QNE Interior node is not able to reserve the number of requested resources, the "PHR_Resource_Request" control information that is associated with the <Bandwidth> parameter MUST be marked. In this situation the RMD specific operation for unsuccessful reservation will be applied (see [Section 4.6.1.2](#)).

- * When the modification request requires a decrease of the reserved resources, the QNE Ingress node MUST include this value into the <Bandwidth> parameter of the "RMD QoS Description" field. Subsequently an RMD release procedure SHOULD be accomplished (see [Section 4.6.1.5](#)).

4.6.1.5 RMD release procedure

If a refresh RESERVE message does not arrive at a QNE Interior node within the refresh time-out period then the resources associated with this message are removed. This soft state behavior provides certain robustness for the system ensuring that unused resources are not reserved for long time. Resources can be removed by explicit release at any time.

When the RMD-RMF of a QNE edge or QNE Interior node processes a "PHR_Release_Request" control information it MUST identify the <PHB-CLASS> parameter and estimate the time period that elapsed after the previous refresh. This MAY be done by indicating the time lag, say "T_lag", between the last sent "PHR_Refresh_Update" and the "PHR_Release_Request" control information container by the QNE Ingress node. The value of "T_Lag" is first normalized to the length of the refresh period, say "T_period". The ratio between the "T_Lag" and the length of the refresh period, "T_period", is calculated. This ratio is then introduced into the <Time Lag> parameter of the "PHR_Release_Request" control information. When a node (QNE edge or QNE Interior) receives the "PHR_Release_Request" control information, it MUST store the arrival time. Then it MUST calculate the time difference, say "Tdiff", between the arrival time and the start of the current refresh period, "T_period". Furthermore, this node MUST derive the value of the "T_Lag", from the <Time Lag> parameter.

This can be found by multiplying the value included in the <Time Lag> parameter 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 <Bandwidth> parameter of the "RMD QoS Description" field that has been sent together with the "PHR_Release_Request" control information container, 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 4.6.1.5.1](#)) or it can be triggered by either an intra-domain RESPONSE or an end-to-end NOTIFY message that includes a marked (i.e., PDR <M> and/or PDR <S> parameters are set ON) "PDR_Reservation_Report" or "PDR_Congestion_Report".

4.6.1.5.1. Triggered by a RESERVE message

This RMD explicit release procedure can be triggered by a tear (TEAR flag set ON) end-to-end RESERVE message. When a tear (TEAR flag set ON) end-to-end RESERVE message arrives to the QNE Ingress then the QNE Ingress node SHOULD process the message in a standard QoS-NSLP way (see [[QoS-NSLP](#)]). In addition to this, the RMD RMF MUST be notified. It will generate an intra-domain RESERVE(RMD-QSPEC) message. Before generating this message, the RMD RMF is using the RMD traffic class (PHR) resources (specified in <Bandwidth>) and the PHB type (specified in <PHB-CLASS>) for a RMD release procedure. This can be achieved by subtracting the amount of the requested resources from the total reserved amount of resources

stored in the RMD traffic class state.

Bader, et al.

[Page 28]

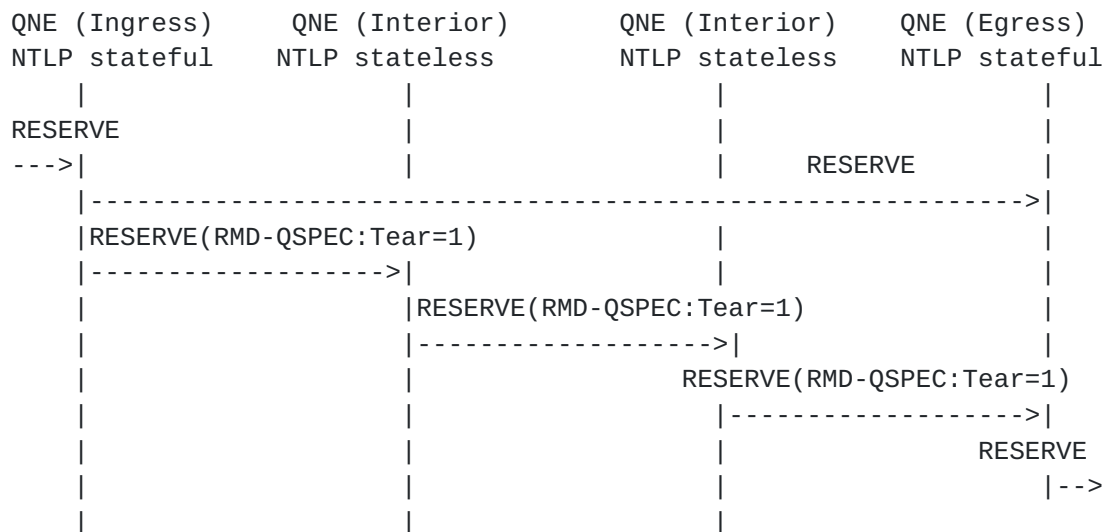


Figure 11: Explicit release triggered by RESERVE used by the RMD-QOSM

The intra-domain RESERVE (RMD-QSPEC) message MUST include a "RMD QoS Description" field and a PHR container, (i.e., "PHR_Resource_Release") and it MAY include a PDR container, (i.e., PDR_Release_Request). An example of this operation can be seen in Figure 11.

Most of the non default values of the objects contained in the tear intra-domain RESERVE message are set by the QNE Ingress node in the same way as described in [Section 4.6.1.1](#). The following objects are set differently:

- * the flag "Acknowledge" (A) SHOULD be set "OFF";
- * The <RII> object is not included in this message. This is because the QNE Ingress node does not need to receive a response from the QNE Egress node;
- * the TEAR flag is set to ON;
- * the PHR resource units MUST be included into the <Bandwidth> parameter of the "RMD QoS Description" field;
- * the value of the <Admitted Hops> parameter has to be set to "1";
- * the value of the <Time Lag> parameter of the PHR container is calculated by the RMD-QOSM functionality (see 4.6.1.5) the value of the <Control Type> parameter of PHR container is set to "3" (i.e., PHR_Resource_Release).

The intra-domain tear RESERVE (RMD-QSPEC) message is received and processed by the QNE Interior nodes. Most of the non-default values of the objects contained in this refresh intra-domain RESERVE (RMD-QSPEC) message are set by a QNE Interior node in the same way as described in [Section 4.6.1.1](#). The following objects are set and processed differently:

- * Any QNE Interior node that receives the combination of the "RMD QoS Description" field and the "PHR_Resource_Release" control information container, it MUST identify the traffic class (PHB) and release the requested resources included in the <Bandwidth> parameter. This can be achieved by subtracting the amount of RMD traffic class requested resources, included in the <Bandwidth> parameter, from the total reserved amount of resources stored in the RMD traffic class state. The value of the <Time Lag> parameter of the "PHR_Resource_Release" container is used during the release procedure as explained in [Section 4.6.1.5](#).

The intra-domain tear RESERVE (RMD-QSPEC) message is received and processed by the QNE Egress node. The "RMD QoS Description" and the "PHR RMD-QOSM control " container (and if available the "PDR RMD-QOSM control information" container) are read and processed by the RMD QoS signaling model functionality. The value of the <Bandwidth> parameter of the "RMD QoS Description" field and the value of the <Time Lag> field of the PHR container 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 <Bandwidth> parameter, 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., QNE Egress) only if the intra-domain tear RESERVE (RMD-QSPEC) message arrives at the QNE Egress node.

[4.6.1.5.2](#) Triggered by a marked RESPONSE or NOTIFY message

This RMD explicit release procedure can be triggered by either an end-to-end RESPONSE message with a <M> marked PDR container (see [Section 4.6.1.2](#)) or an intra-domain NOTIFY (PDR) message (see [Section 4.6.1.6](#)) with a <M> or <S> marked PDR container. This RMD specific release procedure can be terminated at any QNE edge or any QNE Interior node using the <Max_Admitted Hops> field.

The RMD specific explicit release procedure that is terminated at a QNE Interior (or QNE edge) node is denoted as RMD specific partial release procedure. This explicit release procedure can be used, for example, during a RMD specific operation for unsuccessful reservation (see [Section 4.6.1.2](#)) or severe congestion (see [Section 4.6.1.6](#)). When the RMD QoS signaling model functionality of a QNE Ingress node receives a <M> or <S> marked PDR container of type "PDR_Reservation_Report" or "PDR_Congestion_Report", it MUST start an RMD partial release procedure. The QNE Ingress node generates an intra-domain RESERVE (RMD-QSPEC) message. Before generating this message, the RMD-QOSM 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 intra-domain RESERVE (RMD-QSPEC) message is triggered by an intra-domain NOTIFY(PDR) message then the intra-domain RESERVE(RMD-QSPEC) message MUST include a <RMD QoS Description> field and a PHR container, (i.e., PHR_Resource_Release) and it MAY include a PDR container, (i.e., PDR_Release_Request). An example of this message exchange can be seen in Figure 12.

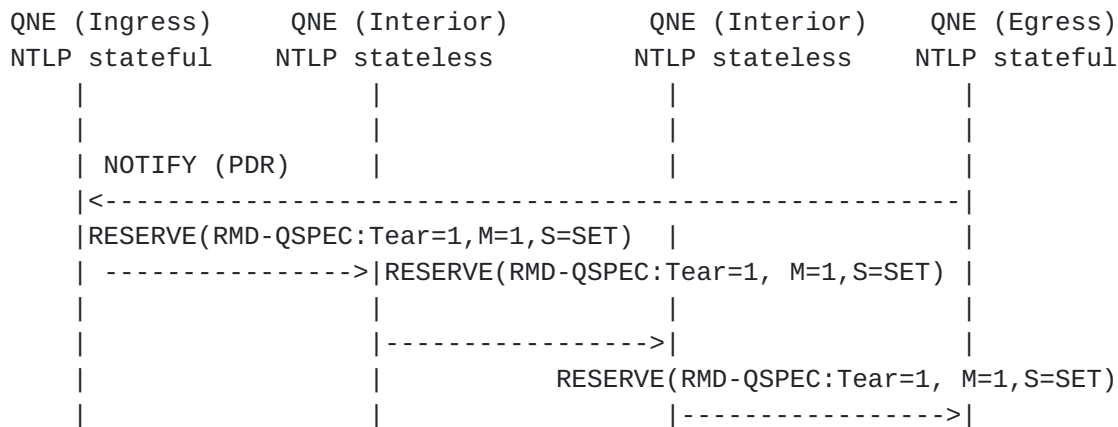


Figure 12: Basic operation during RMD explicit release procedure triggered by NOTIFY used by the RMD-QOSM

When the generation of the intra-domain RESERVE(RMD-QSPEC) message is triggered by an end-to-end RESPONSE(PDR) message then this generated intra-domain RESERVE(RMD-QSPEC) message MUST include a <RMD QoS Description> field and a PDR container, (i.e., PHR_Resource_Release) and it MAY include a PDR container, (i.e., PDR_Release_Request). An example of this operation can be seen in Figure 13.

Most of the non-default values of the objects contained in the tear intra-domain RESERVE(RMD-QSPEC) message are set by the QNE Ingress node in the same way as described in [Section 4.6.1.1](#).

The following objects MUST be used and/or set differently:

- * The value of the <M> parameter of the PHR container MUST be set to "1".
- * When the tear intra-domain RESERVE message is triggered by a NOTIFY message, then the value of the <S> parameter of the PHR container MUST be set to "1".
- * The RESERVE message MAY include PDR container.
- * When the tear intra-domain RESERVE message is triggered by an intra-domain RESPONSE(RMD-QSPEC) message, then the value of the <Max Admitted Hops> parameter of the PDR container included in the received <M> marked intra-domain RESPONSE(PDR) message MUST be included in the <Max Admitted Hops> parameter of the PDR container of the RESERVE message.

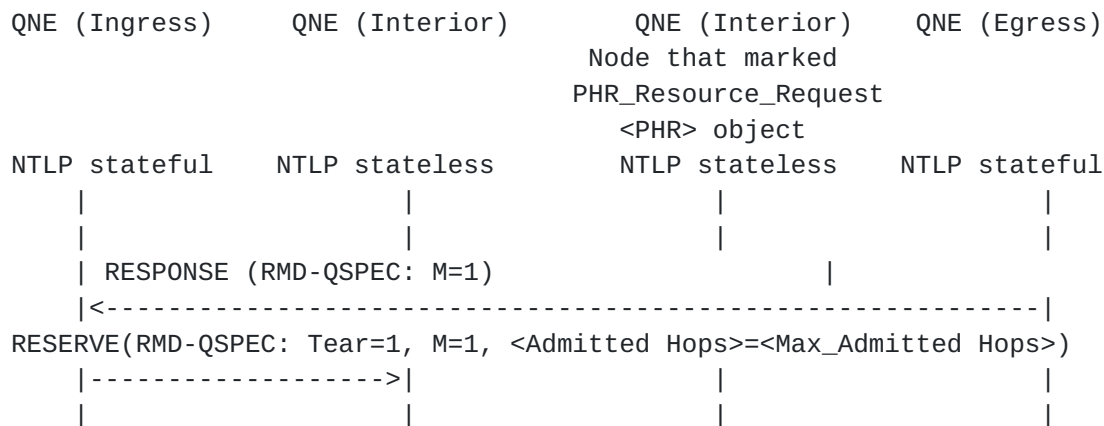


Figure 13: Basic operation during RMD explicit release procedure
Triggered by RESPONSE used by the RMD-QOSM

Any QNE edge or QNE Interior node that receives the "RMD QoS Description" field and the PHR container MUST identify the traffic class state (PHB), using the <PHB-CLASS> parameter, and release the requested resources included in the <Bandwidth> field. This can be achieved by subtracting the amount of RMD traffic class requested resources, included in the <Bandwidth> field, from the total reserved amount of resources stored in the RMD traffic class state. The value of the <Time Lag> parameter of the PHR field is used during the release procedure as explained in [Section 4.6.1.5](#).

The <Admitted Hops> value included in the PHR container is increased by one. If the value of <M> parameter of the "PHR_Resource_Release" control information container is "1" and if the value of the <S> parameter is set to "0" then the <Max_Admitted Hops> value included in the PDR container MUST be compared with the calculated <Admitted Hops> value. When these two values are equal then the intra-domain RESERVE(RMD-QSPEC) has to be terminated and it will not be forwarded downstream. The reason of this is that the QNE node that is currently processing this message was the last QNE node that successfully processed the "RMD QoS Description" field and PHR container of its associated initial reservation request (i.e., initial intra-domain RESERVE(RMD-QSPEC) message). Its next QNE downstream node was unable to successfully process the initial reservation request, therefore, this QNE node marked the <M> parameter of the "PHR_Resource_Request" control information. When the values of the <M> and <S> parameters are set to "0", then this message will not be terminated by a QNE Interior node, but it will be forwarded in the downstream direction. The QNE Egress node will receive and process the PHR_Resource_Release control information. Afterwards, the QNE Egress node MUST terminate the intra-domain RESERVE(RMD-QSPEC) message.

4.6.1.6. Severe congestion handling

This section describes the operation of the RMD-QOSM when a severe congestion occurs within the Diffserv domain.

When a failure in a communication path, e.g. router or link failure occurs, the routing algorithms will adapt to failures by changing the routing decisions to reflect changes in the topology and traffic volume. As a result the re-routed traffic will follow a new path, which may result in overloaded nodes as they need to support more traffic than their capacity allows. This may cause a severe congestion in the communication path. In this situation the available resources, may not be enough to meet the required QoS for all the flows along the new path. Therefore, one or more flows SHOULD be terminated, or forwarded in a lower priority queue.

Interior nodes notify edge nodes by data marking (proportional marking) or marking the refresh messages using the <S> and <Overload %> parameters.

4.6.1.6.1 Severe congestion handling by the RMD-QOSM refresh procedure

The QoS-NSLP and RMD are able to cope with congested situations using the refresh procedure, see [Section 4.6.1.3](#). If the refresh is not successful in an QNE Interior node, edge nodes are notified by "S" marking the refresh messages and by including the percentage of overload into the <Overload %> field in the "PHR_Refresh_Update" container, carried by the intra-domain RESERVE message.

The intra-domain RESPONSE message that is sent by the QNE Egress towards QNE Ingress will contain a PDR container with a Parameter/Container ID = 10, i.e., "PDR_Congestion_Report". The values of the <S> and <Overload %> fields of this container should be set equal to the values of the <S> and <Overload %> fields, respectively, carried by the PHR_Refresh_Update" container. The flows that cannot be supported, i.e., based on the value included in the <Overload %> parameter, are terminated, or forwarded in a lower priority queue. The flows can be terminated by using the RMD release procedure described in [Section 4.6.1.5](#).

In general, relying the soft state refresh mechanism solves the congestion within the time frame of the refresh period. If this mechanism is not fast enough additional functions SHOULD be used, which are described in [Section 4.6.1.6.2](#).

4.6.1.6.2 Severe congestion handling by proportional data packet marking

This severe congestion handling method requires the following additional functionalities.

4.6.1.6.2.1 Operation in the Interior nodes

The QNE Interior node detecting severe congestion marks data packets passing the node in which the severe congestion was detected. For the severe congestion marking, two additional DSCPs SHOULD be allocated for each traffic class. One MAY be used to indicate that the packet passed a congested node. The other DSCP MUST be used to indicate the degree of congestion by marking the bytes proportionally to the degree of congestion. Note however, that it is RECOMMENDED that the total number of additional DSCPs within a RMD domain, needed for severe congestion handling MUST not exceed the limit of 16.

4.6.1.6.2.2 Operation in the Egress nodes

The QNE Egress node applies a predefined policy to solve the severe congestion, by selecting a number of inter domain (end-to-end) flows that SHOULD be terminated, or forwarded in a lower priority queue. For these flows (sessions), the QNE Egress node generates and sends a NOTIFY(PDR) message to the QNE Ingress node (its upstream stateful QoS-NSLP peer) to indicate the severe congestion in the communication path. This message MUST include a PDR container ("PDR_Reservation_Report").

The non-default values of the objects contained in the NOTIFY(PDR) message MUST be set by the QNE Egress node as follows:

- * the values of the <ERROR_SPEC> object is set by the standard QoS-NSLP protocol functions.
- * the value of the Parameter/Container ID of the PDR container SHOULD be set to "10" (i.e., PDR_Congestion_Report).
- * The value of the <M> parameter of the PDR container MUST be set to "1".
- * The value of the <S> parameter of the PDR container MUST be set to "1".

4.6.1.6.2.3 Operation in the Ingress nodes

Upon receiving the (end-to-end) NOTIFY message, the QNE 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 shifting them to an alternative RMD traffic class (PHB). Note that due to the fact that the QoS-NSLP state in the QNE Ingress node maintains the binding between the end-to-end and intra-domain sessions, the QNE Ingress node can associate the PDR container to the right intra-domain session.

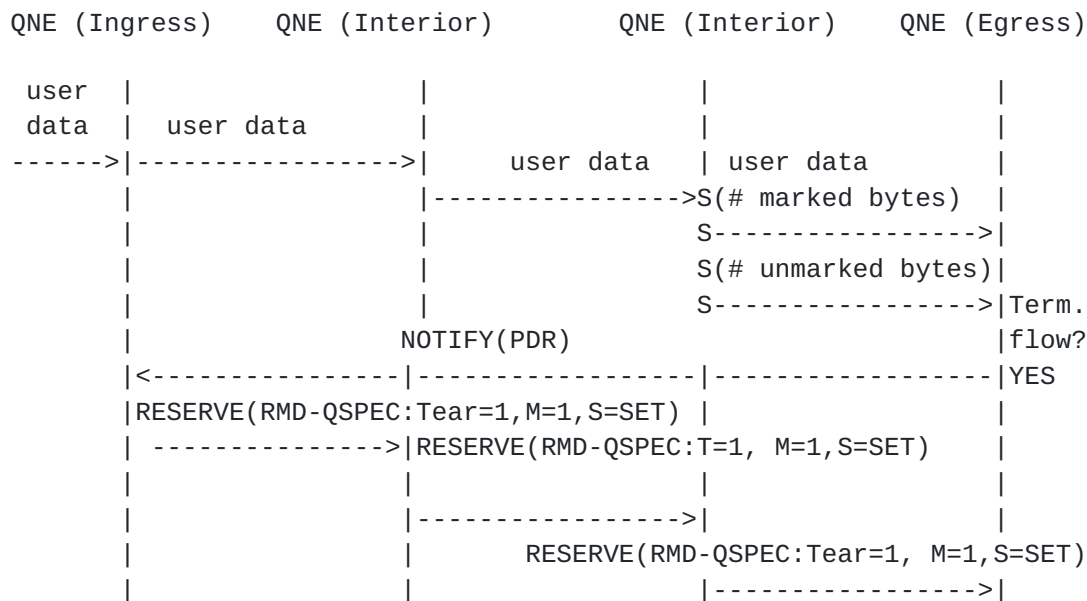


Figure: 14 RMD severe congestion handling

The severe congestion notification function based on data marking can be used for implementing a simple admission control within a Diffserv domain, see Figure 14. In one or a few nodes along the data thresholds are set in the resource management function for the data traffic belonging to different PHBs. If the threshold is exceeded, the data packets are marked in the DSCP field to indicate the high load of different PHBs. In this case the Egress node sends a NOTIFY(PDR) message to the Ingress node, which MAY block the incoming traffic belonging to the same PHB until the traffic volume decreases below the threshold, or forwards it in a lower priority queue.

4.6.2 Bi-directional operation

RMD assumes asymmetric routing by default. Combined sender-receiver initiated reservation cannot be efficiently done in the RMD domain because upstream NTLP states are not stored in Interior routers. Therefore the bi-directional operation SHOULD be performed by two sender-initiated reservations (sender&sender). We assume that the QNE edge nodes are common for both upstream and downstream directions, therefore, the two reservations/sessions can be bound at the QNE edge nodes.

This bi-directional sender&sender procedure can then be applied between the QNE edges (QNE Ingress and QNE Egress) nodes of the RMD QoS signaling model. In the situation that a security association exists between the QNE Ingress and QNE Egress nodes (see Figure 15), and the QNE Ingress node has the required <Bandwidth> parameters for both directions, i.e., QNE Ingress towards QNE Egress and QNE Egress towards QNE Ingress, then the QNE Ingress MAY include both <Bandwidth> parameters (needed for both directions) into the RMD-QSPEC within a RESERVE message. In this way the QNE Egress node is able to use the QoS parameters needed for the "Egress towards Ingress" direction (QoS-2). The QNE Egress is then able to create a RESERVE with the right QoS parameters included in the QSPEC, i.e., RESERVE (QoS-2). Both directions of the flows are bound by inserting the <BOUND_SESSION_ID> object at the QNE Ingress and QNE Egress.

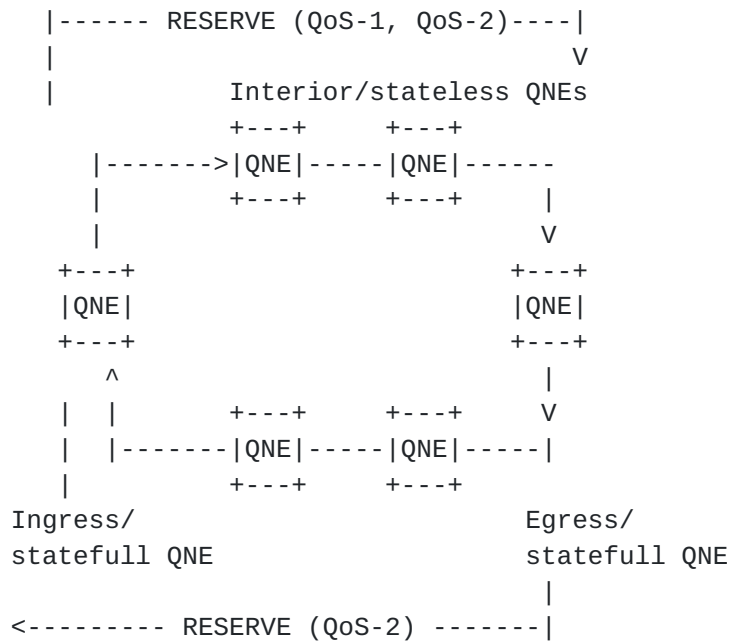


Figure 15: The bi-directional reservation scenario in the RMD domain

A bidirectional reservation, within the RMD domain, is indicated by the PHR and PDR flags, which are set in all messages. Upstream end-to-end messages include the session ID of downstream messages using BOUND_SESSION_ID and vice versa.

If no security association exists between the QNE Ingress and QNE Egress nodes the bi-directional reservation for the sender&sender scenario in the RMD domain SHOULD use the scenario specified in [QoS-NSLP] as "Bi-directional reservation for sender&sender scenario".

In the following sections it is considered that the QNE edge nodes are common for both upstream and downstream directions and therefore, the two reservations/sessions can be bounded at the QNE edge nodes. Furthermore, it is considered that a security association exists between the QNE Ingress and QNE Egress nodes, and the QNE Ingress node has the required <Bandwidth> parameters for both directions, i.e., QNE Ingress towards QNE Egress and QNE Egress towards QNE Ingress.

4.6.2.1 Successful and unsuccessful reservations

This section describes the operation of the RMD-QOSM where a RMD bi-directional reservation operation is either successfully or unsuccessfully accomplished.

The bi-directional successful reservation is similar to a combination of two unidirectional successful reservations that are accomplished in opposite directions, see Figure 16. The main differences of the bi-directional successful reservation procedure with the combination of two unidirectional successful reservations accomplished in opposite directions are as follows. The intra-domain RESERVE message sent by the QNE Ingress node towards the QNE Egress node, is denoted in Figure 16 as RESERVE (RMD-QSPEC): "forward". The main differences between the RESERVE (RMD-QSPEC): "forward" message used for the bi-directional successful reservation procedure and a RESERVE (RMD-QSPEC) message used for the unidirectional successful reservation are as follows:

- * the RII object is not included in the message. This is because no RESPONSE message is expected to arrive.
- * the bit of the PHR container indicates a bi-directional reservation and is set to "1".
- * the PDR container is also included into the RESERVE(RMD-QSPEC): "forward" message. The value of the Parameter/Container ID is "4", i.e., "PDR_Reservation_Request". Note that the response PDR container sent by a QNE Egress to a QNE Ingress node is not carried by an end-to-end RESPONSE message, but it is carried by an intra-domain RESERVE message that is sent by the QNE Egress node towards the QNE Ingress node (denoted in Figure 16 as RESERVE(RMD-QSPEC):"reverse").
- * the PDR bit indicates a bi-directional reservation and is set to "1".
- * the <PDR Reverse Requested Resources> field specifies the requested bandwidth that has to be used by the QNE Egress node to initiate another intra-domain RESERVE message in the reverse

direction.

Bader, et al.

[Page 38]

The RESERVE(RMD-QSPEC):"reverse" message is initiated by the QNE Egress node at the moment that the RESERVE(RMD-QSPEC):"forward" message is successfully processed by the QNE Egress node. The main differences between the RESERVE(RMD-QSPEC):"reverse" message used for the bi-directional successful reservation procedure and a RESERVE(RMD-QSPEC) message used for the unidirectional successful reservation are as follows:

- * the RII object is not included in the message. This is because no RESPONSE message is expected to arrive;
- * the value of the <Bandwidth> parameter is set equal to the value of the <PDR Reverse Requested Resources> field included in the RESERVE(RMD-QSPEC):"forward" message that triggered the generation of this RESERVE(RMD-QSPEC): "reverse" message;
- * the value of the [BOUND_SESSION_ID] object is set equal to the SESSION_ID of the intra domain session associated with the RESERVE(RMD-QSPEC):"forward" message that triggered the generation of this RESERVE(RMD-QSPEC):"reverse" message;
- * the bit of the PHR container indicates a bi-directional reservation and is set to "1";
- * the PDR container is included into the RESERVE(RMD-QSPEC):"reverse" message. The value of the Parameter/Container ID is "7", i.e., "PDR_Reservation_Report";
- * the PDR bit indicates a bi-directional reservation and is set to "1".

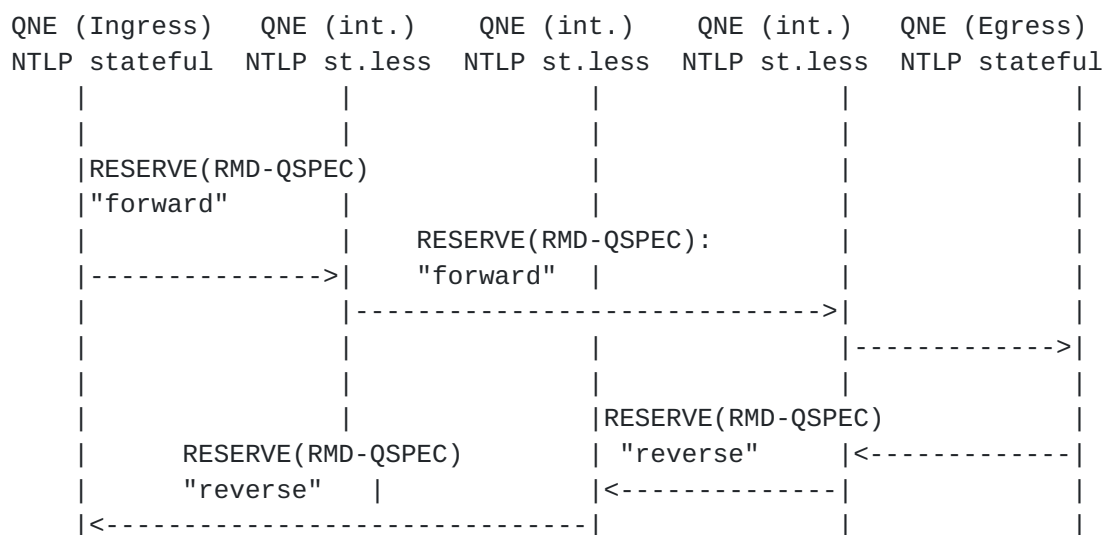


Figure 16: Intra-domain signaling operation for successful bi-directional reservation

Figure 17 and Figure 18 show the flow diagrams used in case of a unsuccessful bi-directional reservation. In Figure 17 it is considered that the QNE that is not able to support the requested <Bandwidth> is located in the direction QNE Ingress towards QNE Egress. In Figure 18 it is considered that the QNE that is not able to support the requested <Bandwidth> is located in the direction QNE Egress towards QNE Ingress.

The main differences between the bi-directional unsuccessful procedure shown in Figure 17 and the bi-directional successful procedure are as follows:

- * the QNE node that is not able to reserve resources for a certain request is located in the "forward" path, i.e., path from QNE Ingress towards the QNE Egress.
- * the QNE node that is not able to support the requested <Bandwidth> it MUST mark the <M> bit, i.e., set to value "1", of the RESERVE(RMD-QSPEC): "forward".

The operation for this type of unsuccessful bi-directional reservation is similar to the operation for unsuccessful uni-directional reservation shown in Figure 9. The main difference is that the QNE Egress generates an intra-domain (local) RESPONSE(PDR) message that is sent towards QNE Ingress node.

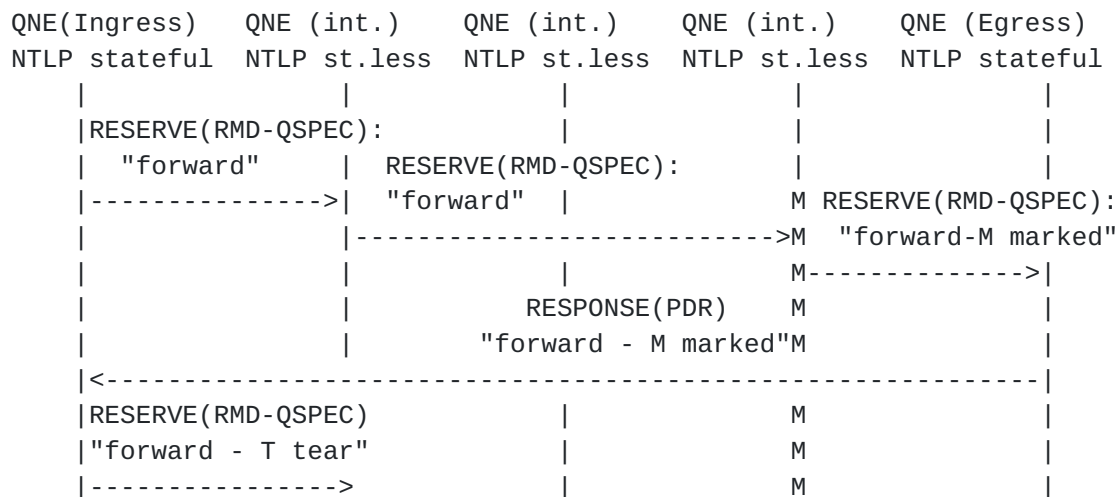


Figure 17: Intra-domain signaling operation for unsuccessful bi-directional reservation (rejection on path QNE(Ingress) towards QNE(Egress))

The main differences between the bi-directional unsuccessful procedure shown in Figure 18 and the in bi-directional successful procedure are as follows:

- * the QNE node that is not able to reserve resources for a certain request is located in the "reverse" path, i.e., path from QNE Egress towards the QNE Ingress.

- * the QNE node that is not able to support the requested <Bandwidth> it MUST mark the <M> bit, i.e., set to value "1", the RESERVE(RMD-QSPEC):"reverse".
- * the QNE Ingress uses the information contained in the received PHR and PDR containers of the RESERVE(RMD-QSPEC): "reverse" and generates a tear intra-domain (local) RESERVE(RMD-QSPEC): "forward - T tear" message. This message carries a "PHR_Release_Request" and a "PDR_Release_Request" control information. This message is sent to QNE Egress node. The QNE Egress node by using the information contained in the "PHR_Release_Request" and the "PDR_Release_Request" control info containers it generates a RESERVE(RMD-QSPEC):"reverse - T tear" message that is sent towards the QNE Ingress node.

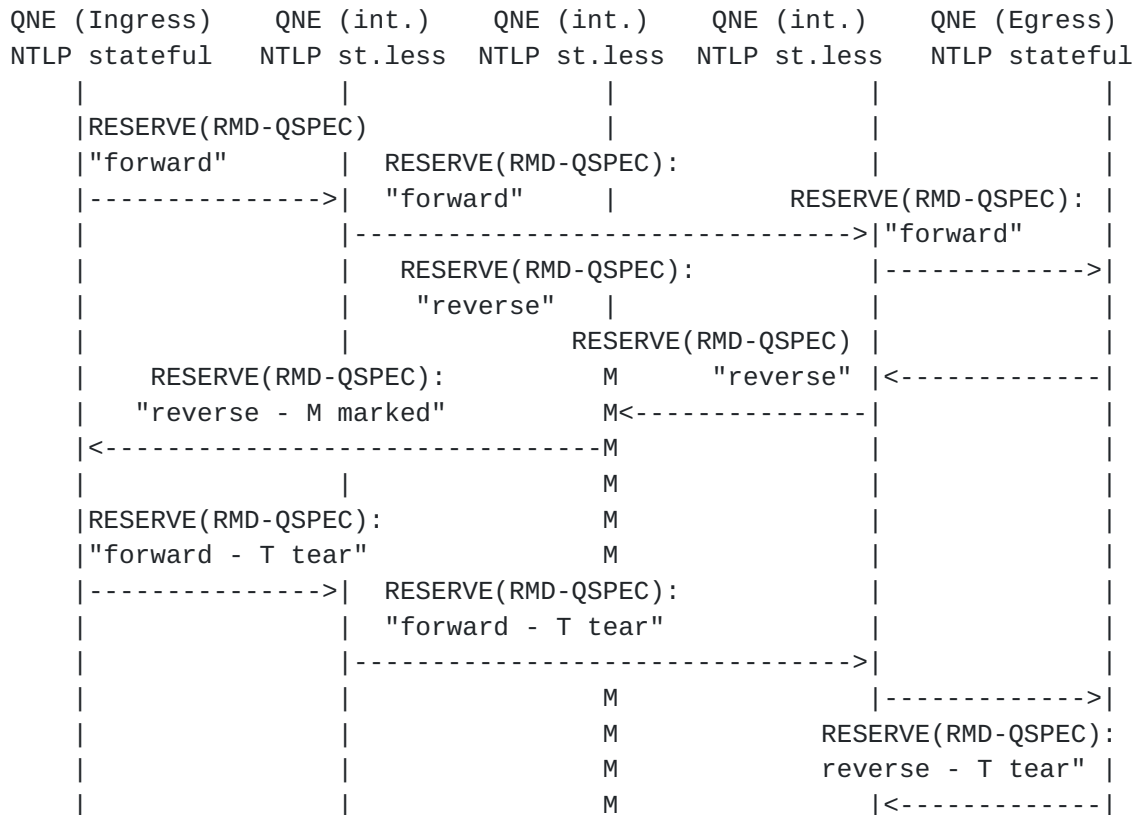


Figure 18: Intra-domain signaling normal operation for unsuccessful bi-directional reservation (rejection on path QNE(Egress) towards QNE(Ingress))

4.6.2.2 Refresh reservations

This section describes the operation of the RMD-QOSM where a RMD bi-directional refresh reservation operation is accomplished.

The refresh procedure in case of RMD reservation-based method follows a similar scheme as the successful reservation procedure, described in [Section 4.6.2.1](#), and depicted in Figure 16 and the way of how the refresh process of the reserved resources is maintained, is similar to the refresh process used for the intra-domain uni-directional reservations (see [Section 4.6.1.3](#)).

Note that the RMD traffic class refresh periods used by the bound bi-directional sessions MUST be equal in all QNE edge and QNE Interior nodes.

The main differences between the RESERVE(RMD-QSPEC):"forward" message used for the bi-directional refresh procedure and a RESERVE(RMD-QSPEC):"forward" message used for the bi-directional successful reservation procedure are as follows:

- * the value of the Parameter/Container ID of the PHR container is "2", i.e., "PHR_Refresh_Update".
- * the value of the Parameter/Container ID of the PDR container is "5", i.e., "PDR_Refresh_Request".

The main differences between the RESERVE(RMD-QSPEC):"reverse" message used for the bi-directional refresh procedure and the RESERVE(RMD-QSPEC): "reverse" message used for the bi-directional successful reservation procedure are as follows:

- * the value of the Parameter/Container ID of the PHR container is "2", i.e., "PHR_Refresh_Update".
- * the value of the Parameter/Container ID of the PDR container is "8", i.e., "PDR_Refresh_Report".

[4.6.2.3](#) Modification of aggregated reservations

This section describes the operation of the RMD-QOSM where a RMD

In the case when the QNE edges maintain, for the RMD QoS model, QoS-NSLP aggregated reservation states and if such an aggregated reservation has to be modified (see [Section 4.3.1](#)) then similar procedures to [Section 4.6.1.4](#) are applied. In particular:

* When the modification request requires an increase of the reserved resources, the QNE Ingress node MUST include the corresponding value into the <Bandwidth> parameter of the "RMD QoS Description" field, which is sent together with a "PHR_Resource_Request" control information. If a QNE edge or QNE Interior node is not able to reserve the number of requested resources, then the "PHR_Resource_Request" control information associated with the <Bandwidth> parameter MUST be marked. In this situation the RMD specific operation for unsuccessful reservation will be applied (see [Section 4.6.2.1](#)).

* When the modification request requires a decrease of the reserved resources, the QNE Ingress node MUST include this value into the <Bandwidth> parameter of the "RMD QoS Description" field. Subsequently an RMD release procedure SHOULD be accomplished (see [Section 4.6.2.4](#)).

[4.6.2.4](#) Release procedure

This section describes the operation of the RMD-QOSM where a RMD bi-directional reservation release operation is accomplished. The message sequence diagram used in this procedure is similar to the one used by the successful reservation procedures, described in [Section 4.6.2.1](#), and depicted in Figure 16. However, the way of how the release of the reservation is accomplished, is similar to the RMD release procedure used for the intra-domain uni-directional reservations (see [Section 4.6.1.5](#) and Figure 17 and Figure 18).

The main differences between the RESERVE (RMD-QSPEC):

"forward" message used for the bi-directional release procedure

and a RESERVE (RMD-QSPEC): "forward" message used for the bi-directional successful reservation procedure are as follows:

- * the value of the Parameter/Container ID of the PHR container is "3", i.e. "PHR_Release_Request";
- * the value of the Parameter/Container ID of the PDR container is "6", i.e., "PDR_Release_Request";

The main differences between the RESERVE (RMD-QSPEC): "reverse" message used for the bi-directional release procedure and the RESERVE (RMD-QSPEC): "reverse" message used for the bi-directional successful reservation procedure are as follows:

- * the value of the Parameter/Container ID of the PHR container is "3", i.e., "PHR_Release_Request";
- * the PDR container is not included in the RESERVE (RMD-QSPEC): "reverse" message.

4.6.2.5 Severe congestion handling

This section describes the severe congestion handling operation used in combination with bi-directional reservation procedures.

This severe congestion handling operation is similar to the one described in [Section 4.6.1.6](#).

4.6.2.5.1 Severe congestion handling by the RMD-QOSM bi-directional refresh procedure

This procedure is similar to the severe congestion handling procedure described in [Section 4.6.1.6.1](#). The difference is related to how the refresh procedure is accomplished, see [Section 4.6.2.2](#) and to how the flows are terminated, see [Section 4.6.2.4](#).

4.6.2.5.2 Severe congestion handling by proportional data packet marking

This section describes the severe congestion handling by proportional data packet marking when this is combined with a bi-directional reservation procedure.

This procedure is similar to the severe congestion handling procedure described in [Section 4.6.1.6.2](#). The main difference is related to the location of the severe congested node, i.e., "forward" path (i.e., path between QNE Ingress towards QNE Egress) or "reverse" path (i.e., path between QNE Egress towards QNE Ingress).

Figure 19 shows the scenario where the severe congested node is located in the "forward" path. This scenario is very similar to the severe congestion handling scenario described in [Section 4.6.1.6.2](#) and shown in Figure 14. The difference is related to the release procedure, which is accomplished in the same way as described in [Section 4.6.2.4](#).

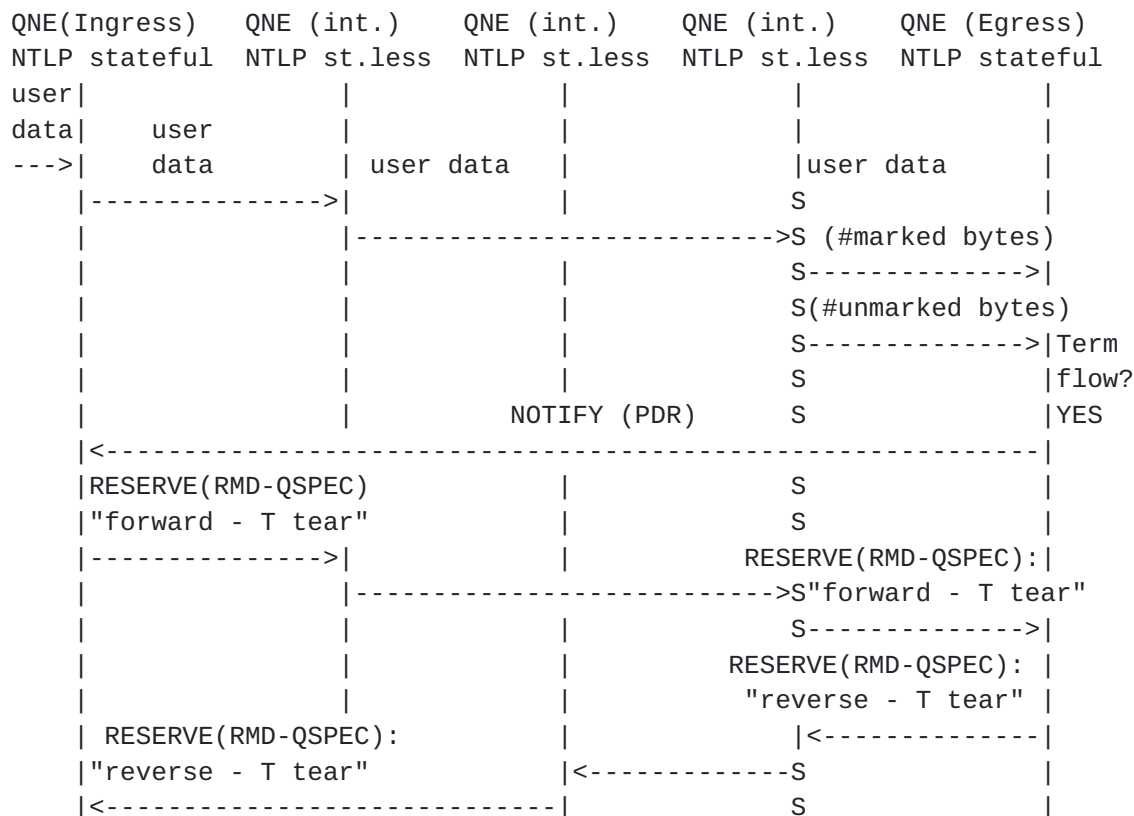


Figure 19: Intra-domain RMD severe congestion handling for bi-directional reservation (congestion on path QNE(Ingress) towards QNE(Egress))

Figure 20 shows the scenario where the severe congested node is located in the "reverse" path. The main difference between this scenario and the scenario shown in Figure 19 is that no intra-domain NOTIFY(PDR) message has to be generated by the QNE Egress node. This is because the (#marked and #unmarked) user data is arriving at the QNE Ingress. The QNE Ingress node will be able to calculate the number of flows that have to be terminated or forwarded in a lower priority queue.

For the flows that have to be terminated a release procedure, see [Section 4.6.2.4](#), is initiated to release the reserved resources on the "forward" and "reverse" paths.

The RMD QOSM aims to be very lightweight signaling with regard to the number of signaling message roundtrips and the amount of state established at involved signaling nodes with and without reduced state on QNEs. This implies the usage of the Datagram Mode which cannot benefit from security protection. As such, RMD signaling is target towards intra-domain signaling only. Still it is possible to provide some degree of security.

In the context of RMD QOSM signaling a classification between in-path adversaries and off-path adversaries needs to be made. Furthermore, it might be necessary to differentiate between always off-path nodes and nodes which are only off-path with regard to a specific signaling message.

The following paragraph aims to raise a discussion about the requirements placed on the security properties of the signaling message exchange:

First, it is necessary to protect the message communication between the QNE Ingress and the QNE Egress. This is possible since these nodes are meant to be stateful nodes and do not suffer from the same constraints as network QNE Interior nodes. This mechanism already ensures that intermediate or off-path nodes initiate some signaling messages towards the edges. An adversary is therefore unable to inject an NOTIFY message or a RESERVE message. Additionally, such a security protection ensures that only selected fields can be modified. To accomplish this type of protection two mechanisms need to be considered that both require enhancements to the QoS NSLP. Since the intra-domain RESERVE message travels along several stateless nodes it is necessary to provide a protection at the QoS-NSLP. Channel security at the GIMPS layer might in most cases not be possible due to the nature of the NTLP datagram mode message. One option is the usage of the Cryptographic Message Syntax (CMS) to protect selected payloads at the QoS NSLP layer. A digital signature is suitable if the QNE Ingress and the QNE Egress node do not need to share a secret nor do they require an in-band exchange of certificates due to the closed environment where a pre-distribution of certificates can be assumed. Such a digital signature would amount for about roughly 600 to 700 bytes of payloads within a packet. Further implementation experience will be required to see whether this message size is within the MTU limits for the entire NSIS message. The usage of a digital signature for a one-shot packet would, however, allow an adversary located within the intra-domain network to flood the QNE Ingress or QNE Egress with digitally signed messages. This would require heavy computation by the target nodes and could lead to a denial of service. The usage of an out-of-band authentication and key exchange protocol extending the Internet Key Exchange Protocol using a Domain of Interpretation is a good alternative. An example of this approach was exercised in [[RSVP-DOI](#)].

The QNE Ingress node should know its QNE Egress node based on e.g. an end-to-end signaling communication. In the reverse direction routing state has already been established as part of GIMPS signaling.

The congestion handling mechanism is very difficult to detect since the malicious behavior might be hard to distinguish from regular behavior. Hence, intrusion detection techniques and statistical measurements could help to detect a malicious node within the RMD aware network domain. This technique has been suggested also for DiffServ Codepoint packet marking (add ref. later). A general observation can be made here that a router implementing a QoS signaling protocol (and the RMD QOSM) can, similar to a router without support for QoS signaling, do a lot of harm to a system.

6. IANA Considerations

RMD-QOSM requires a new IANA registry.

7. Open issues

This section describes the open issues related to the RMD QoS signaling model. More details on open issues will be provided in a future version of this draft.

7.1 Explicit congestion notification

Explicit congestion notification (ECN) described in [RFC 3168](#) might be used to complement RMD basic functions. Congestion notification can be based on queue management, e.g. RED.

8. Acknowledgments

The authors express their acknowledgement to people who have worked on the RMD concept: Z. Turanyi, R. Szabo, A. Csaszar, A. Takacs, G. Pongracz, A. Marquetant, O. Pop, V. Rexhepi, D. Partain, M. Jacobsson, S. Oosthoek, P. Wallentin, P. Goering, A. Stienstra, M. de Kogel, M. Zoumaro-djayoon, M. Swanink.

9. Authors' Addresses

Attila Bader
Traffic Lab
Ericsson Research
Ericsson Hungary Ltd.
Laborc 1
Budapest, Hungary, H-1037
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

Cornelia Kappler
Siemens AG
Siemensdamm 62
Berlin 13627, Germany
Email: cornelia.kappler@siemens.com

Hannes Tschofenig
Siemens AG
Otto-Hahn-Ring 6
Munich 81739, Germany
EMail: Hannes.Tschofenig@siemens.com

Tom Phelan
Sonus Networks
250 Apollo Dr.
Chelmsford, MA USA 01824
EMail: tphelan@sonusnet.com

10. Normative References

[QoS-NSLP] Bosch, S., Karagiannis, G. and A. McDonald, "NSLP for Quality-of-Service signaling", [draft-ietf-nsis-qos-nslp-05](#) (work in progress), July 2005.

[QSP-T] Ash, J., Bader, A., Kappler C., "QoS-NSLP QSpec Template" [draft-ietf-nsis-QSpec-04](#) (work in progress), July 2005.

11. Informative References

[RFC2205] Braden, R., Zhang, L., Berson, S., Herzog, A., Jamin, S., "Resource ReSerVation Protocol (RSVP)-- Version 1 Functional Specification", IETF [RFC 2205](#), 1997.

[RFC2961] Berger, L., Gan, D., Swallow, G., Pan, P., Tommasi, F. and S. Molendini, "RSVP Refresh Overhead Reduction Extensions", [RFC 2961](#), April 2001.

[RFC3175] Baker, F., Iturralde, C. Le Faucher, F., Davie, B., "Aggregation of RSVP for IPv4 and IPv6 Reservations", IETF [RFC 3175](#), 2001.

[GIMPS] Schulzrinne, H., Hancock, R., "GIMPS: General Internet Messaging Protocol for Signaling", [draft-ietf-nsis-ntlp-04](#) (work in progress), Oct 2004.

[RFC1633] Braden R., Clark D., Shenker S., "Integrated Services in the Internet Architecture: an Overview", [RFC 1633](#)

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z. and W. Weiss, "An Architecture for Differentiated Services", [RFC 2475](#), December 1998

[RFC2638] Nichols K., Jacobson V., Zhang L. "A Two-bit Differentiated Services Architecture for the Internet", [RFC 2638](#), July 1999

[RMD1] Westberg, L., et al., "Resource Management in Diffserv (RMD): A Functionality and Performance Behavior Overview", IFIP PFHSN'02

[RMD2] G. Karagiannis, et al., "RMD - a lightweight application of NSIS" Networks 2004, Vienna, Austria.

[RMD3] Marquetant A., Pop O., Szabo R., Dinnyes G., Turanyi Z., "Novel Enhancements to Load Control - A Soft-State, Lightweight Admission Control Protocol", Proceedings of the 2nd International Workshop on Quality of future Internet Services, Coimbra, Portugal, Sept 24-26, 2001, pp. 82-96.

[RMD4] A. Csaszar et al., "Severe congestion handling with resource management in diffserv on demand", Networking 2002

[RSVP-DOI] Tschofenig H., Schulzrinne H., "RSVP Domain of Interpretation for ISAKMP ", [draft-tschofenig-rsvp-doi-00.txt](#), (work in progress), May 2003

12. Intellectual Property Statement

IPR Statement about RMD

I hereby give the following IPR Disclosure in relation to the RMD concept proposed by Ericsson and currently under discussion in IEFT WG NSIS:

To the best of my knowledge there are no Ericsson patents or filed patent applications on RMD protocol operation or basic principles. To my knowledge there is only one Ericsson patent application family that could possibly be relevant merely to particular implementation of RMD. This patent family comprises US patent 6687655 and counterparts in other countries.

To the best of my knowledge there is only one Ericsson owned invention without any patent applications filed yet that could possibly be relevant to particular implementation of RMD, but this invention is not relevant to RMD protocol operation or basic principles.

I have been authorized by Ericsson to give the following Licensing Declaration in relation to the RMD concept proposed by Ericsson and discussed in IEFT WG NSIS:

In case a license to a patent in the patent family above or a patent issued/granted on an application for patent on the invention above should be necessary for implementing any Internet Standard, Ericsson is willing to grant to anybody a license to such patent on fair, reasonable and non-discriminatory conditions for the implementation of the standard, subject to reciprocity.

Attila Bader

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in [BCP 78](#) and [BCP 79](#).

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Copyright (C) The Internet Society (2005).

This document is subject to the rights, licenses and restrictions contained in [BCP 78](#), and except as set forth therein, the authors retain all their rights.