International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Y.2122

(06/2009)

SERIES Y: GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS AND NEXT-GENERATION NETWORKS

Next Generation Networks – Quality of Service and performance

Flow aggregate information exchange functions in NGN

Recommendation ITU-T Y.2122

ITU-T Y-SERIES RECOMMENDATIONS

GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS AND NEXT-GENERATION NETWORKS

GLOBAL INFORMATION INFRASTRUCTURE	
General	Y.100–Y.199
Services, applications and middleware	Y.200–Y.299
Network aspects	Y.300–Y.399
Interfaces and protocols	Y.400–Y.499
Numbering, addressing and naming	Y.500–Y.599
Operation, administration and maintenance	Y.600–Y.699
Security	Y.700–Y.799
Performances	Y.800–Y.899
INTERNET PROTOCOL ASPECTS	
General	Y.1000–Y.1099
Services and applications	Y.1100–Y.1199
Architecture, access, network capabilities and resource management	Y.1200–Y.1299
Transport	Y.1300–Y.1399
Interworking	Y.1400–Y.1499
Quality of service and network performance	Y.1500–Y.1599
Signalling	Y.1600–Y.1699
Operation, administration and maintenance	Y.1700–Y.1799
Charging	Y.1800–Y.1899
IPTV over NGN	Y.1900–Y.1999
NEXT GENERATION NETWORKS	
Frameworks and functional architecture models	Y.2000–Y.2099
Quality of Service and performance	**Y.2100–Y.2199**
Service aspects: Service capabilities and service architecture	Y.2200–Y.2249
Service aspects: Interoperability of services and networks in NGN	Y.2250–Y.2299
Numbering, naming and addressing	Y.2300–Y.2399
Network management	Y.2400–Y.2499
Network control architectures and protocols	Y.2500–Y.2599
Future networks	Y.2600–Y.2699
Security	Y.2700–Y.2799
Generalized mobility	Y.2800–Y.2899
Carrier grade open environment	Y.2900–Y.2999

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T Y.2122

Flow aggregate information exchange functions in NGN

Summary

Recommendation ITU-T Y.2122 specifies the requirements and architecture for flow aggregate information exchange functions (FIXF). FIXF allows network domains to exchange flow aggregate information that reflects the current network conditions. In order to guarantee end-to-end QoS in NGN, cooperation among networks in an end-to-end path is necessary. Exchange of network configuration information is a key component of cooperation among participating networks. Flow aggregation has been considered a scalable technique for traffic management. Additional information obtained through FIXF can improve the accuracy of performance estimation and effectiveness of resource management.

Source

Recommendation ITU-T Y.2122 was approved on 29 June 2009 by ITU-T Study Group 13 (2009-2012) under Recommendation ITU-T A.8 procedures.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

© ITU 2010

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

CONTENTS

Page

Introduction

This Recommendation specifies the requirements and architecture for flow aggregate information exchange functions (FIXF). It addresses the flow or flow aggregate (FA) performance prediction by way of exchanging the network operating parameters or unique parameters of flow aggregates. There are a few critical parameters (e.g., number of hops, link capacity, etc.) for the performance of a flow for a given QoS architecture, which are sometimes tuned dynamically (from time-to-time, from flow-to-flow) within a network. By having the other networks' parameters for the flow (or for the flow aggregate to which the flow belongs), one can infer the performance of the flow in these networks and in its own network. The estimation can be complemented by measurement, i.e., the estimation rules can be further adjusted by the measured performances. Consider the following example: A delay bound for an expedited forwarding (EF) flow in a differentiated services (DiffServ) network can be calculated by using parameters given from neighbouring networks. The initial delay estimation rule can be just to adopt the delay bound obtained from the calculation. The bound, however, can be too conservative, compared to the measured delay, especially when the maximum burst sizes of flows in the same class are large. Then one can adjust the estimation rule to reduce the expected delay value under condition that the maximum burst sizes are large.

One obvious benefit of this approach is that upon a flow admission request, network operators can immediately decide whether the requested end-to-end performances can be met, based on the estimation rules that have been adjusted and stabilized. The frequency of the exchange, estimation, comparison, and adjustment can vary according to implementation. The frequency of exchange for a given flow aggregate, for example, depends on the parameter value precision, flow granularity, intended level of accuracy, etc.

The overall process for the performance estimation is depicted in Figure Intro.1.

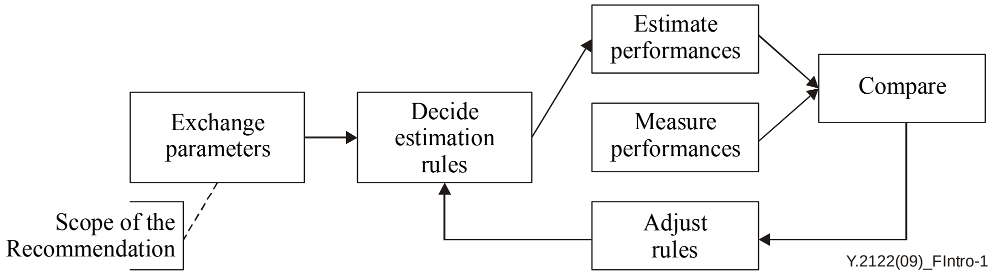

Figure Intro.1 – Estimation process example based on exchanged parameters

The network operating parameters to be exchanged vary according to network QoS architecture. For example, in the DiffServ [b-IETF RFC 2475] architecture (or 802.1Q, 802.11e-type networks) the important parameters include the behaviour aggregate (BA) to which a packet belongs, and the bandwidth utilization of the BA based on the implied admission policy (e.g., not more than 10% for EF BA). In another example, the flow aggregation granularity and the number of hops in an aggregation region, in case of a flow-based network such as flow state aware network [b-ITU-T Y.2121], are the important parameters.

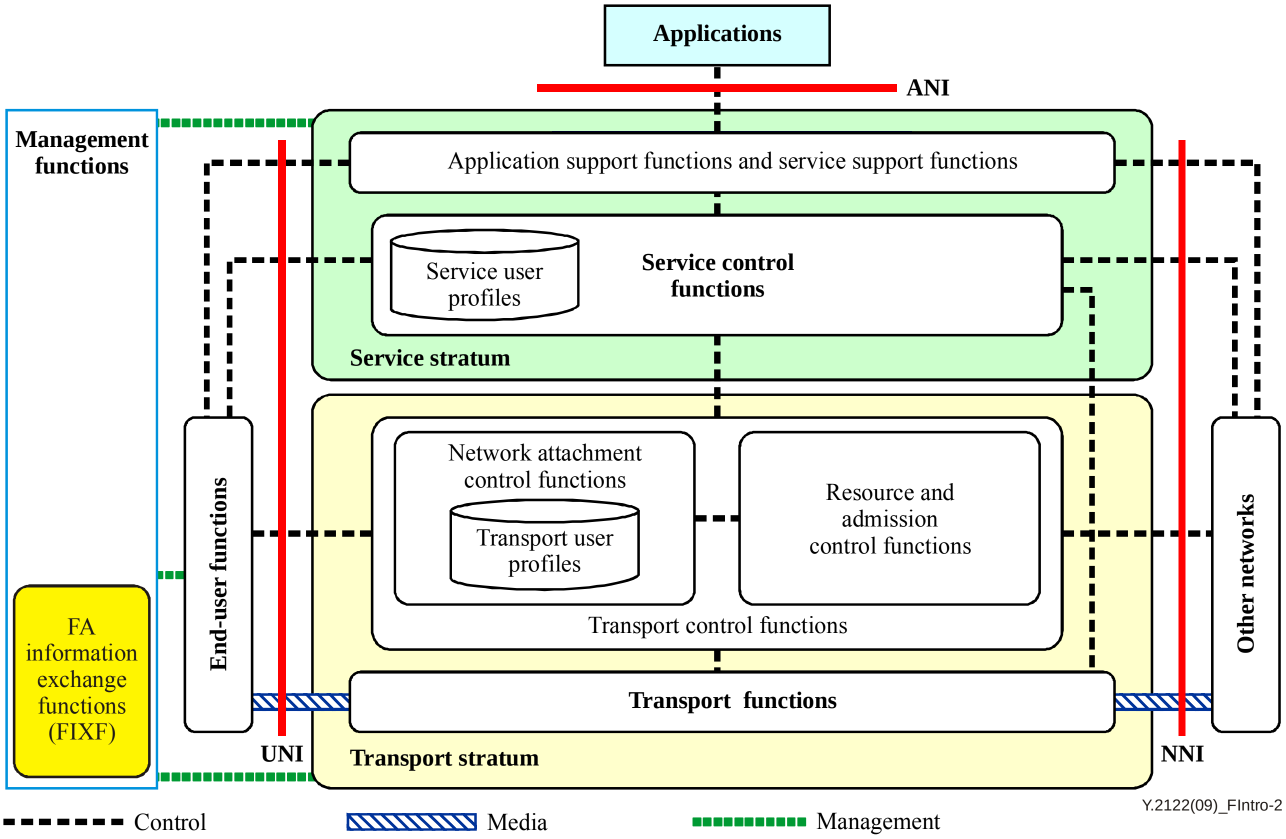
Applications
ANI
Management functions
Application support functions and service support functions
Service user profiles
Service control functions
Service stratum
Network attachment control functions
Transport user profiles
Resource and admission control functions
Transport control functions
End-user functions
Other networks
Transport functions
FA information exchange functions (FIXF)
UNI
Transport stratum
NNI
Y.2122(09)_FIntro-2
Control
Media
Management

Figure Intro.2 – Position of the FA information exchange functions in the NGN architecture [ITU-T Y.2012]

As depicted in Figure Intro.2, the FA information exchange functions defined in this Recommendation are one of the NGN management functions. The goal of the functions is to provide additional information to improve the accuracy of the admission decision. The FIXF collects the FA configuration information throughout the networks. The collected information can be used for estimating the end-to-end QoS, thus possible applications of this function can be the resource and admission control functions (RACF) or other management functions that handle end-to-end QoS.

Recommendation ITU-T Y.2122

Flow aggregate information exchange functions in NGN

1 Scope

This Recommendation defines the network operating parameters of a flow aggregate (FA) to be exchanged among different network domains for various types of FAs in NGN, and defines the architecture for exchanging these parameters. It also defines, for informational purposes, the mapping of the transport technology specific parameters to the defined network operating parameters. The details of the protocols are for further study and are out of the scope of this Recommendation. Usage examples of the exchanged operating parameters are shown in several appendices.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T Y.2012] Recommendation ITU-T Y.2012 (2006), *Functional requirements and architecture of the NGN release 1.*

[ITU-T Y.2701] Recommendation ITU-T Y.2701 (2007), *Security requirements for NGN release 1.*

3 Definitions

This Recommendation defines the following terms:

3.1 aggregation end point: The node where a flow aggregate is de-aggregated. Aggregate start and end points for a flow aggregate can collocate in the same node.

3.2 aggregation region: A set of neighbouring nodes within which every flow maintains its membership of a flow aggregate unaltered. For example, in a DiffServ network, the aggregation region is usually limited to a single node, since at the very next node the membership within a behaviour aggregate is likely to be changed. An aggregation region is defined per flow aggregate. Typically, a node can belong to multiple aggregation regions at the same time.

3.3 aggregation start point: The node where flows are aggregated to form a flow aggregate.

3.4 flow: A set of IP packets that have the same IP 5-tuple in case of IPv4 packets, or have the same flow label in case of IPv6 packets.

3.5 flow aggregate: A set of flows treated in the same way in terms of scheduling.

3.6 flow aggregate membership: The flows within the flow aggregate. The membership can be identified by the set of flow IDs within the flow aggregate. A given flow at a given node cannot be a member of different flow aggregates at the same time.

3.7 flow aggregate table: A set of flow aggregate table entries.

3.8 flow aggregate table entry: A set of information for a flow aggregate including flow aggregate specific configuration parameters. A flow aggregate table entry is the elementary unit to be maintained and exchanged.

3.9 number of hops: The number of nodes in an aggregation region, including both aggregation start point and aggregation end point.

3.10 policy region: A set of neighbouring nodes within which a single policy regarding the flow aggregation is applied. In this Recommendation, it is assumed that an administrative domain is a single policy region. A "network" or a "domain" means a policy region.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AEP	Aggregation End Point
AR	Aggregation Region
ASP	Aggregation Start Point
BC	Bandwidth Constraint
Bs	Sustainable Burst Size
CT	Class Type
DiffServ	Differentiated Services
DSCP	DiffServ Code Point
DS-MPLS	DiffServ over MPLS
ECN	Explicit Congestion Notification
FA	Flow Aggregate
FIE-FE	FA Information Exchange functional entity
FIG-FE	FA Information Gathering functional entity
FIM-FE	FA Information Management functional entity
FIXF	Flow Aggregate Information Exchange Functions
FSA	Flow State Aware
IntServ	Integrated Services
LSB	Least Significant Bit
LSP	Label Switched Path
MAM	Maximum Allocation Model
MSB	Most Significant Bit
PHB	Per Hop Behaviour
R	Representative
RACF	Resource and Admission Control Functions
RDM	Russian Dolls Model
Rs	Sustainable Transfer Rate
TE	Traffic Engineering
WFQ	Weighted Fair Queuing

5 Conventions

None.

6 High level requirements for FA information exchange functions

The FA information exchange functions described in this Recommendation are required to:

1) be able to gather information on flow aggregates configuration parameters from networks, which are physically separated and are willing to exchange their information;
2) be able to provide gathered information to application entities which can utilize the global FA information to facilitate end-to-end QoS;
3) have minimum impact on the performance of the underlying transport layer; and
4) be flexible in updating, exchanging, and gathering transaction's frequencies and triggering methods.

7 Information exchange architecture

This clause provides:

– a description of functional entities of the FA information exchange functions; and

– a description of modes of the FA information exchange functions required to establish the management view of a complete end-to-end customer service and the hierarchy of management information flows required to provide and maintain such a service.

It should be noted that the modes of information exchange presented here are lists of reasonable approaches to meet the requirements for orderly provision of end-to-end QoS services which should allow operators to cooperate; or an operator maintaining multiple domains to cooperate among domains, in support of this. More modes of operation can further be defined.

7.1 Overall architecture

Figure 7-1 describes the functional architecture of FA information exchange with the functional entities. (For a detailed description of the functional entities, see clause 7.4). This architecture includes:

– FIM-FE (FA information management functional entity): FIM-FE is responsible for maintaining and updating the information on a per-flow aggregate basis. This information is captured in a form of a flow aggregate table.

– FIE-FE (FA information exchange functional entity): FIE-FE is responsible for providing the FA table to the FIG-FE.

– FIG-FE (FA information gathering functional entity): FIG-FE is the final point of the information gathering and processing. FIG-FE is responsible for handling requests from applications it is associated with.

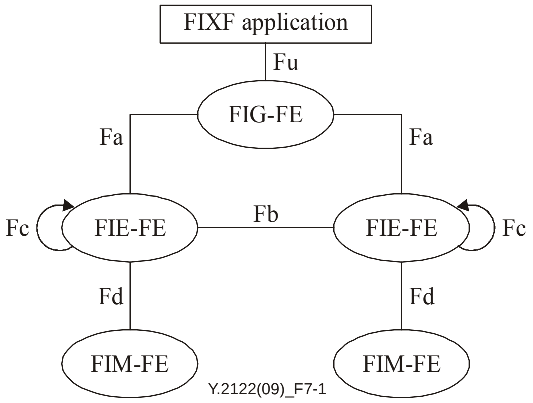

Figure 7-1 – Functional architecture for FA information exchange

7.2 Modes of operation

The management of operations for information exchange, collecting and processing over multiple network domains is a challenging issue. Recently, the exchange procedures for the QoS measurement information over networks have been standardized both in ITU-T [b-ITU-T Y.2173] and IETF [b-IETF RFC 5470]. This clause defines the operation modes for generic flow aggregate information exchange based on the distributed models such as specified in [b-ITU-T Y.2173]. A network participating in flow aggregate information exchange functions is required to support the following modes of operations.

7.2.1 Hub mode

In Hub mode, it is assumed that a central entity exists for cooperation among networks in the end-to-end path. The central entity governs all the transactions (information exchanges) among networks. Note that this does not imply there should only be one central entity over Internet.

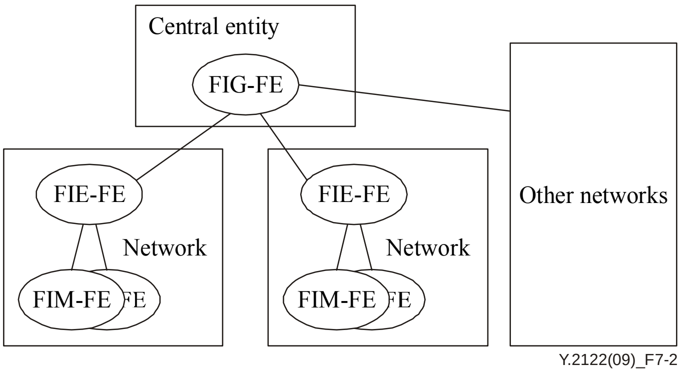

Figure 7-2 – Hub mode operation

7.2.2 Star mode

In Star mode, the central entity described in Hub mode is distributed to edge networks, usually in the ingress edge networks (by an edge network it is meant a network that directly interfaces to an access network or CPN). The central entity communicates with downstream or upstream networks. One of the most obvious advantages of the Star mode over the hub mode is its scalability.

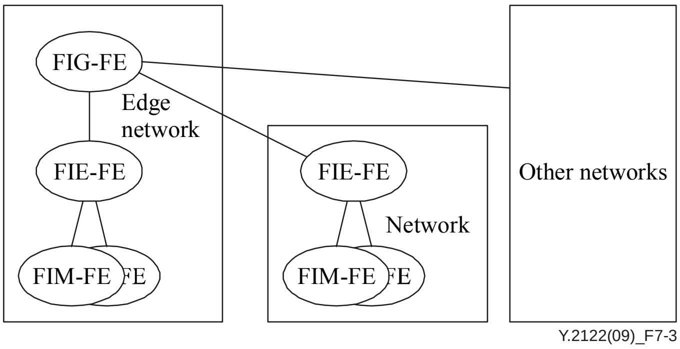

Figure 7-3 – Star mode operation

7.2.3 Cascade mode

Cascade mode can be seen as a peer-to-peer mode. Networks exchange information with only directly neighbouring networks.

It is expected that the use of the Star and the Hub modes when there are many interconnecting network domains would be increasingly difficult to manage and vulnerable to configuration delays or errors. Smaller operators would have to offer exchange services and for them the use of the Star or Hub mode would be inappropriate in many instances. Such operators would probably consider using the Cascade mode, depicted below. Use of the Star or Hub mode requires contractual agreements between the service providers and each network domain. In summary, the use of a single point of contact for the customer, as in the Hub or Star mode and a single point of management control for the service instance is an attractive feature of these modes. By contrast, use of the Cascade mode makes it possible to build exchange services on a global basis while only maintaining contractual relationships with adjacent domains. Thus, the mode is more scalable than the Star and Hub modes and may be more appropriate for use by 'non-global' operators.

A limitation of the Cascade mode is that it gives the service initiator less management control of the whole IP service path. However, with a well designed interface between operators, the service initiator will have as much information and control of the end-to-end service as the Star and Hub modes give.

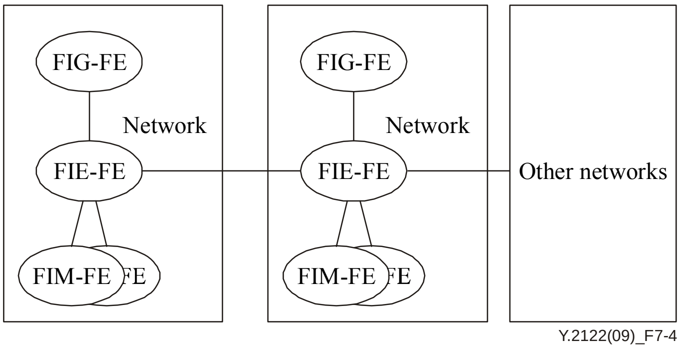

Figure 7-4 – Cascade mode operation

7.2.4 Hybrid mode

Any combination of hub, Star, and Cascade modes can be considered as Hybrid mode.

7.3 Operation of FA information exchange functions

This clause explains the procedure of the FA information collection. A typical network topology including physical locations of the functional entities is described in Appendix I. The following information components are exchanged over the reference points:

- Requestor identifier: a unique identifier for the sender of the request message.
- Responder identifier: a unique identifier for the sender of the response message.
- Reporter identifier: a unique identifier for the sender of the report message.
- FA identifier: a unique identifier for a FA (e.g., MPLS LSP).
- FA parameters: the list of FA parameters specified in Table 2.

Table 1 summarizes the exchanged messages and associated information components.

Table 1 – Messages for the FA parameter exchange

Exchanged message	Information components	Reference points	Direction
FA-information-request	Requestor identifier	Fa	FIG-FE to FIE-FE
	FA identifiers (optional)	Fb and Fc	FIE-FE to FIE-FE
		Fd	FIE-FE to FIM-FE
FA-information-response	Responder identifier	Fa	FIE-FE to FIG-FE
	FA identifiers (optional)	Fb and Fc	FIE-FE to FIE-FE
	FA parameters	Fd	FIM-FE to FIE-FE
FA-information-report	Reporter identifier	Fa	FIE-FE to FIG-FE
	FA identifiers (optional)	Fb and Fc	FIE-FE to FIE-FE
	FA parameters	Fd	FIM-FE to FIE-FE

The operation of the FA information collection is performed either by request or report. FA-information-request and FA-information-response are used in the request model, and FA-information-report message is used in the report model. The FEs form three levels of hierarchy. In the request model, the higher level FE requests the FA information to the lower level FE. In the report model, the lower level FE sends the FA configuration changes to the higher level FE.

The request operation is initiated by FIG-FE. The request operation does not require the real-time level dynamics. The response time of the request can be several seconds. The information collection occurs periodically or on an on-demand basis. In a periodical request operation, the polling interval can vary according to the dynamics of the network. In the normal case, it can be in the order of several minutes. Figure 7-5 shows the FA request procedure. The FIG-FE polls the FIE-FE for the FA information of the domain. The FIE-FE, then sends the response to the requester FIG-FE with the updated FA information of domain. When a domain contains multiple aggregation regions, one FIE-FE represents the domain and provides the domain information to the FIG-FE. FIE-FEs other than the representing FIE-FE exchange the information through the representing FIE-FE. If the FA information of FIE-FE is not the latest one, FIE-FE may send the FA-information-request to the FIM-FE to collect the latest information. It is recommended that the polling response contain the FA parameters of all the FAs under the responder's control. It is optional that the requestor specify the FA identifier to collect the FA parameter of the specified FA.

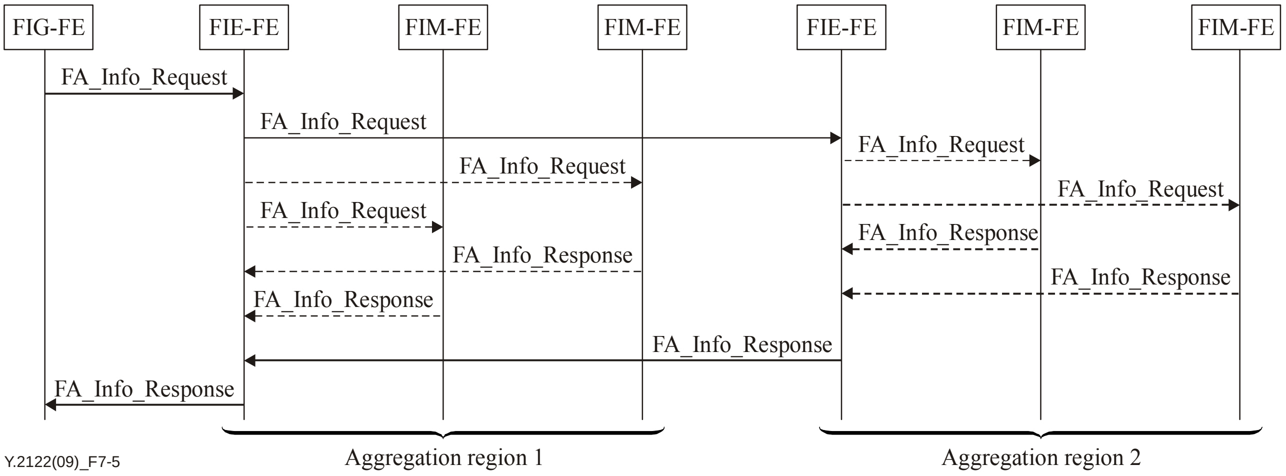

Figure 7-5 – Procedure for the request operation for FA information

In the report operation, the report is sent from the lower level FE to the higher level FE. The report can be sent periodically or whenever the FA configuration is changed. It is recommended that the FIM-FE send a report when the local FA information is changed. Figure 7-6 shows the procedure of FA information exchange in the report model. FIM-FE sends the report to FIE-FE whenever the local FA configuration is changed. Once FIE-FE receives the information from the FIM-FE, the FIE-FE composites the received information and sends the domain level report to FIG-FE. If the configuration change is significant, FIE-FE can send the report to FIG-FE immediately after receiving a report from FIM-FE. Otherwise, FIE-FE stores the updated FA information and sends it at the next reporting period.

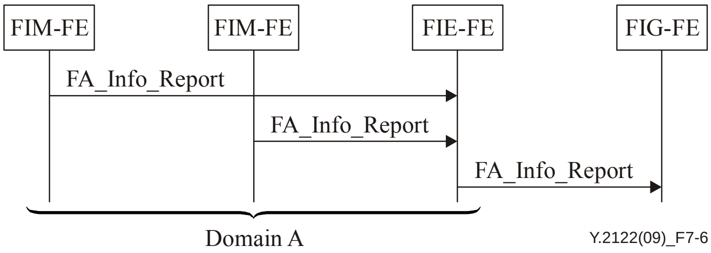

Figure 7-6 – Procedure for the report operation for FA information update

In real implementations, the request and report models are all used together. For example, the reference point Fd between FIE-FE and FIM-FE may work in the report model while Fa between FIE-FE and FIG-FE works in the request model. In this case, the FIE-FE receives the FA parameters whenever the configuration change occurs in FIM-FE. The updated FA information is collected by FIG-FE periodically.

7.4 Functional entities

7.4.1 FA information management functional entity (FIM-FE)

FIM-FE is specific to transport technology (e.g., MPLS). FIM-FE is responsible for maintaining and updating the information on a per-flow aggregate basis, by which the scheduling and queuing are done in a network. This information is captured in a form of flow aggregate table. Ensuring the scalability is the key reason for per-FA basis handling of information. It is required to have at least one FIM-FE in an aggregation region. Multiple aggregation regions may share a single FIM-FE, for example if they share a node with the co-located FIM-FE. FIM-FE resides optionally in either ASP (aggregation start point) or AEP (aggregation end point). Example use cases are described in Appendix IV. Typically, there are multiple FIM-FEs within a network, for example, one per each ASP. Note that many flow aggregates share an ASP (or an AEP). An FIM-FE is recommended to manage those FAs. FIM-FE is required to provide the following functions:

– FA parameters mapping: FIM-FE maps the transport-technology specific FA parameters and network operation parameters into the parameters defined in this Recommendation.

– FA table entry creation: Upon a flow admission, FIM-FE creates and populates an entry in the FA table if the flow does not belong to any existing FA. The specific mechanism for obtaining values of the entry is out of the scope of this Recommendation.

– FA table update: Upon an event such as a flow admission, a flow release, or a flow resource modification, FIM-FE updates the corresponding FA table entry. Not every event is required to trigger an update.

– FA table entry deletion: Upon a last flow release, FIM-FE deletes the corresponding entry from the FA table.

– FA table content delivery: Upon request from FIE-FE, the requested part of the FA table is read and sent to FIE-FE.

– FA table status report: Optionally, FIM-FE may trigger an interruption to notify that the FA table is updated, deleted, or created.

In order to be able to create and update the FA table entry, FIM-FE is required to be able to gather the necessary information from appropriate sources, including the policy server, edge switches, or other transport stratum entities.

7.4.2 FA information exchange functional entity (FIE-FE)

FIE-FE is responsible for providing the FA table to FIG-FE. The request from FIG-FE can be either dynamic or subscription-based. FIE-FE is required to be able to support both cases. In the dynamic case, FIE-FE sends FA table to FIG-FE upon receiving request from FIG-FE. In the subscription-based case, FIE-FE sends the FA table to FIG-FE upon receiving the update from FIM-FE.

In order to respond to the request from FIG-FE, FIE-FE is required to be able to locate the sources of information, collect information from associated FIM-FEs or other FIE-FEs of different networks or of different aggregation regions, and deliver information to the requesting FIG-FE.

In order to process the report from FIM-FE, FIE-FE is required to be able to locate the destinations of the information, and deliver information to FIG-FEs or other FIE-FEs of different networks or of different aggregation regions; and deliver information to the subscribing FIG-FE.

FIE-FE is required to be able to exchange information on a per-flow aggregate basis, on a per-network basis, or on any granularity in-between.

FIE-FE is independent of transport technology. At least one FIE-FE is required to reside within a single physical network. Assuming network edge nodes are the ASPs and AEPs for all flow aggregates, one possible implementation practice is to put an FIM-FE and FIE-FE pair into every edge node of a network. Another possibility is to put an FIM-FE into every edge node and a single FIE-FE in a network. Allowing multiple FIE-FEs in a network ensures scalability. Example use cases are described in Appendix V.

7.4.3 FA information gathering functional entity (FIG-FE)

FIG-FE is the final point of the information gathering and processing. FIG-FE is required to reside in a network in an end-to-end path. FIG-FE is responsible for handling requests from applications it is associated with. Requests from applications are generally related to the end-to-end QoS provisioning. In order to meet the request from applications, FIG-FE is required to be able to gather FA tables, process them, and generate the relevant information that can be used by the requesting application. Examples of how applications utilize information from FIG-FEs are described in Appendices II and III.

7.5 Reference points

7.5.1 Reference point Fa

The reference point Fa allows the FA information gathering functional entity (FIG-FE) to gather information from FA information exchange functional entities (FIE-FEs) from various networks, including the FIE-FEs of other domains. Fa also allows the requests from FIG-FE to be transferred to FIE-FE.

7.5.2 Reference point Fb

The reference point Fb allows the exchange of information among FIE-FEs in different domains. In Cascade mode, where FIG-FE is fully distributed, the information gathering from remote domains is done through the cooperation among FIE-FEs of multiple domains. In such a case, domain level multi-hop exchange is necessary. The Fb also allows the domain level multi-hop exchange.

7.5.3 Reference point Fc

The reference point Fc allows the exchange of information among FIE-FEs within a domain, in the case where there are multiple FIE-FEs in a single domain.

7.5.4 Reference point Fd

The reference point Fd allows the uploading of the managed information to FIE-FE from FIM-FE. Fd also allows the requests from FIE-FE to be transferred to FIM-FE.

7.5.5 Reference point Fu

The reference point Fu allows the information gathered and processed by FIG-FE to be transferred to applications. Fu also allows the requests from applications to be transferred to FIG-FE.

8 Information to be exchanged

The following parameters are required to be gathered, stored, updated, and exchanged over the networks. A complete set of the listed parameters is a FA table entry. A possible allocation scheme of bits and values for the parameters is described in Appendix VI. Also, since the listed parameters' corresponding terminologies and concepts are different among networks, example mappings for the standard parameters and network-specific parameters are described in Appendix VII.

8.1 Parameters not specific to a FA

8.1.1 Mapping rule from flow ID to FA ID

This field specifies the transport technology of the network and more specifically the policy (or the rule) with which the network maps a flow to a proper flow aggregate.

It is required that this field be exchanged per network and not be stored and exchanged per FA.

Networks are to exchange information on whether their unique policy for mapping IP-level flows into a FA is static or dynamic. By informing such a rule, the informed network can infer to which FA the flow under estimation would belong. Sometimes, it is enough to notify the transport technology, but sometimes it is not. How to describe exactly the complex nature of the mapping policy is an open question.

8.1.2 Bandwidth allocation enforcement method or scheduling method

This field specifies the scheduling method employed by the network, with which the network can provision a certain amount of bandwidth resource to a flow aggregate.

It is required that this field be exchanged per network and not be stored and exchanged per FA.

It is clear that the scheduling policy determines the bandwidth allocated to a given FA, explicitly or implicitly. The schedulers can be categorized into the WFQ (weighted fair queuing)-type rate-guaranteeing schedulers (the explicit rate enforcers), the priority schedulers (the implicit ones), and others. The FIFO and strict priority schedulers can be seen as the special types of the priority schedulers. The schedulers can also be divided into FA-(or flow) based and class-(or service context) based. While the rate-guaranteeing schedulers usually are FA based, and the class-based one is the priority scheduler, the other combinations exist as well. The DiffServ-TE rate allocation methods (Russian Dolls Models [b-IETF RFC 4127], etc.) per class are the examples.

8.2 FA parameters

8.2.1 Representative (R)

This field, R field in short, has a value of either true or false. When true, it means that this set of FA parameters is the representative of an entire network domain, not just for a FA.

A FA table entry is required to reside in every aggregation region for every FA. A FA table, a set of FA table entries, is managed by an FIM-FE. In DiffServ, this could mean that every node has a FA table for BAs (behaviour aggregates). In order to reduce such a complexity, a representative FA table can be maintained per network. Instead of maintaining and exchanging tables for the FAs, a network can optionally maintain a single representative FA table. For example, a DiffServ network can maintain a representative FA table, and send this information to every requesting network.

8.2.2 FA identifier

This field specifies the identity of FA.

This field is void if R field is true.

The value of this field is specific to the transport technology. For example, in an MPLS network, it can be an LSP label.

8.2.3 Number of flows within the FA

This field specifies the current number of flows in a FA. If R field is true, then this field specifies the maximum number of flows in a FA among all the FAs in the network. Because the number of flows in a FA is different from FA to FA, it is required to describe a worst case as a representative of a whole network.

8.2.4 Maximum packet length divided by link capacity

This field specifies the maximum value of packet length, among packets that pass through a link, divided by the link capacity among all the links in the path of the FA in the network.

If R field is true, then this field specifies the maximum value of packet length divided by link capacity across all the links in the network.

It is likely that the capacities of links within an aggregation region vary. While it would be desirable if every capacity is known, it is hard to handle all the information. The minimum of link capacities gives the worst delay bound; therefore this field gives a conservative estimate of delay performance.

8.2.5 Number of hops within the aggregation region or the policy region

This field specifies the number of hops the FA traverses within the aggregation region.

If R field is true, then this field specifies the maximum number of hops within the domain, over a path of any flow. The radius of an aggregation region is a critical parameter for estimating the delay performance of the packets inside a FA. Usually more hops in an aggregation region gives a better delay performance. In a pure DiffServ network, the value of this parameter is one, i.e., the flows within a class are de-aggregated at the very next node.

8.2.6 Maximum number of aggregation regions

If R field is true, then this field specifies the maximum number of aggregation regions a flow traverses in the network, among all the flows in the network. If R field is false, then this field is required to have the value of one.

8.2.7 Maximum sum of sustainable transfer rates (Rs) divided by link capacity

This field specifies the maximum value of the sum of sustainable transfer rates (Rs) of flows within the FA divided by link capacity, across all the links the FA traverses in the aggregation region.

If R field is true, then this field specifies the maximum value of the sum of sustainable transfer rates of flows within a FA divided by link capacity, across all the FAs and all the links the FA traverses in the network. Some QoS architectures do not require the sustainable transfer rate to be specified by end users. In this case this value may be left unknown.

8.2.8 Maximum sum of sustainable burst sizes (Bs) divided by link capacity

This field specifies the maximum value of the sum of sustainable burst sizes (Bs) of flows within the FA divided by link capacity, across all the links the FA traverses in the aggregation region.

If R field is true, then this field specifies the maximum value of sum of sustainable burst sizes of flows within a FA divided by link capacity, across all the FAs and all the links the FA traverses in the network. Some QoS architectures do not require the burst tolerance to be specified by end users. In this case this value may be left unknown.

8.2.9 Minimum guaranteed service rate divided by sum of Rs (if rate-based) or scheduling priority

The type of scheduler used in a network is specified by the field 'Bandwidth allocation enforcement method'. Based on the scheduler type, this field has a different meaning as follows.

If the schedulers in the aggregation region are rate-guaranteeing ones, then this field specifies the minimum value of guaranteed service rate for the FA divided by the sum of Rs of flows within the FA, across all the links within the path the FA traverses.

If the schedulers are priority-based ones, then this field specifies the priority level given to the FA.

In this case, this field represents the relative priority assigned to the FA. How to exactly assign the eight bits into scheduling priorities is technology dependent.

The decision process for the allocated rate, whether it is static or dynamic with measurement, is not covered in this Recommendation.

8.2.10 Packet discard priority

This field specifies the packet discard priority assigned to the FA.

This field corresponds to the drop precedence field in DSCP. It effectively indicates the buffer space allocation priority among FAs.

8.2.11 Class (service context)

This field specifies the service class assigned to the FA.

This field can be mapped to the service context in FSA networks. Each DSCP value for BA indication in DiffServ and each value for the service context indication (e.g., Decimal 10 for AF11 BA) should be used intact.

8.3 Parameters exchanged at the reference points

Table 2 specifies what information parameters are required to be exchanged by which reference points.

Table 2 – Parameters to be exchanged at reference points

Taxonomy	Parameter	Fa	Fb	Fc	Fd
Parameters not specific to a FA	Mapping rule (policy) from Flow ID (IP 5-tuple) to FA ID.	✓	✓		
	Bandwidth allocation enforcement method or (scheduling method).	✓	✓		
FA parameters	Representative (R).	✓	✓	✓	✓
	FA identifier.	✓	✓	✓	✓
	Number of flows within the FA.	✓	✓	✓	✓
	Maximum packet length divided by link capacity.	✓	✓	✓	✓
	Number of hops within the aggregation region or the policy region.	✓	✓	✓	✓
	Maximum number of aggregation regions.	✓	✓	✓	✓
	Maximum sum of sustainable transfer rates (Rs) divided by link capacity.	✓	✓	✓	✓
	Maximum sum of sustainable burst sizes (Bs) divided by link capacity.	✓	✓	✓	✓
	Minimum guaranteed service rate divided by sum of Rs (if rate-based) or scheduling priority.	✓	✓	✓	✓
	Packet discard priority.	✓	✓	✓	✓
	Class (service context).	✓	✓	✓	✓

9 Security considerations

This clause describes security threats and potential attacks and defines security requirements for performing information exchange. The security requirements are based on [ITU-T Y.2701].

Although information exchange among network domains is not expected to add specific security extensions, greater considerations are required in terms of how to establish a trusted model across network domain boundaries. Networks are required to have a means to authenticate, allow, and possibly deny inter-network requests. Also, networks are required to be protected from DoS attacks.

Since the information exchange functions and the use of such information completely rely on other networks' cooperation, if the information from other networks is not trustworthy, then the whole operation would be meaningless. Furthermore, the network's operating status would be revealed if confidentiality is not guaranteed. [ITU-T Y.2701] defines the network elements of the "untrusted zone" and possible threats in the peering network trust model. The following are the threats that may occur in the information exchange procedure.

– alteration of information: This threat includes the destruction, corruption, modification, removal, or loss of information. Possible consequences are false information regarding resource allocation in other networks, inefficient allocation of resources, and misjudgement on admission decisions;

– theft or disclosure of information: This can take place because of the interception of the FA tables or because of granting access to an illegitimate application. This threat may result in various privacy violations;

- interruption of exchange services: This threat is typically realized through a denial of service (DoS) attack. Specifically, the resources (including the computing resources) can be exhausted by forcing them to process too many requests, or by authorization of illegitimate requests. A few known DoS attacks involve: 1) replaying the information exchange request (or response) messages; 2) injection or modification of the request (or response) messages; and 3) flooding, where an adversary sends a large number of requests. The processing of such requests may exhaust the resources, rendering them unavailable for requests from the legitimate operators.

A number of well-known security mechanisms have been either proven or deemed appropriate for mutual authentication and provision of integrity and confidentiality. Transport layer security (TLS) [b-IETF RFC 2246] and IPsec [b-IETF RFC 4301] protocols already employ such mechanisms to provision the transport and network layer security, respectively.

The major security requirements for networks adopting FA information exchange functions are:

- protection of the signalling request/response exchange in support of information exchange;
- protection of the information contained in all functional entities involved in this exchange;
- ensuring the availability and overall expected performance of the information exchange;
- preventing illegitimate access to functional entities from any untrusted network element.

The ways of preventing illegitimate access may include authentication between functional entities from different trust domains, non-repudiation support, firewall, and secure channel mechanisms such as TLS and IPsec.

Appendix I

Typical network topology for FA parameter exchange

(This appendix does not form an integral part of this Recommendation)

The FA parameter exchange procedure is not specific to the network topology. It is required that the control procedure be applicable to any type of network topology. In this appendix, we will use an example of the generic network topology to describe the FA parameter exchange operation.

Figure I.1 describes the generic network topology. The network topology assumes several network domains where each domain of the network makes use of a specific FA technology (e.g., MPLS). For collecting the FA information, the three FEs (e.g., FIM-FE, FIE-FE, and FIG-FE) are implemented in the network. As indicated in Figure I.1, the FIM-FE, FIE-FE, and FIG-FE are structured in a hierarchical manner. The three FEs are implemented in the switch or in the stand-alone server. The detailed description is as follows.

– FIM-FE: The FIM-FE is usually co-located in the edge switches and maintains the information of the local FAs which are aggregated in the switch. Multiple FIM-FEs exist in each aggregate domain.

– FIE-FE: The FIE-FE is responsible for collecting the FA information from multiple FIM-FEs. At least one FIE-FE exists in each aggregate domain. If the network size is big, it is recommended that multiple FIE-FEs exist in the network domain. In Figure I.1, domain C has two FIE-FEs. The multiple FIE-FEs in the same network domains are connected with Fc. It is recommended that only one of the FIE-FE in a domain be connected to the FIG-FE by Fa and to the FIE-FE of other domains by Fb. The FIE-FE is implemented in a separate server or co-located in an edge switch.

– FIG-FE: The FIG-FE is responsible for collecting the FA information from multiple FIE-FEs. There is one FIG-FE in the network. Typically, the FIG-FE is implemented in a separate server.

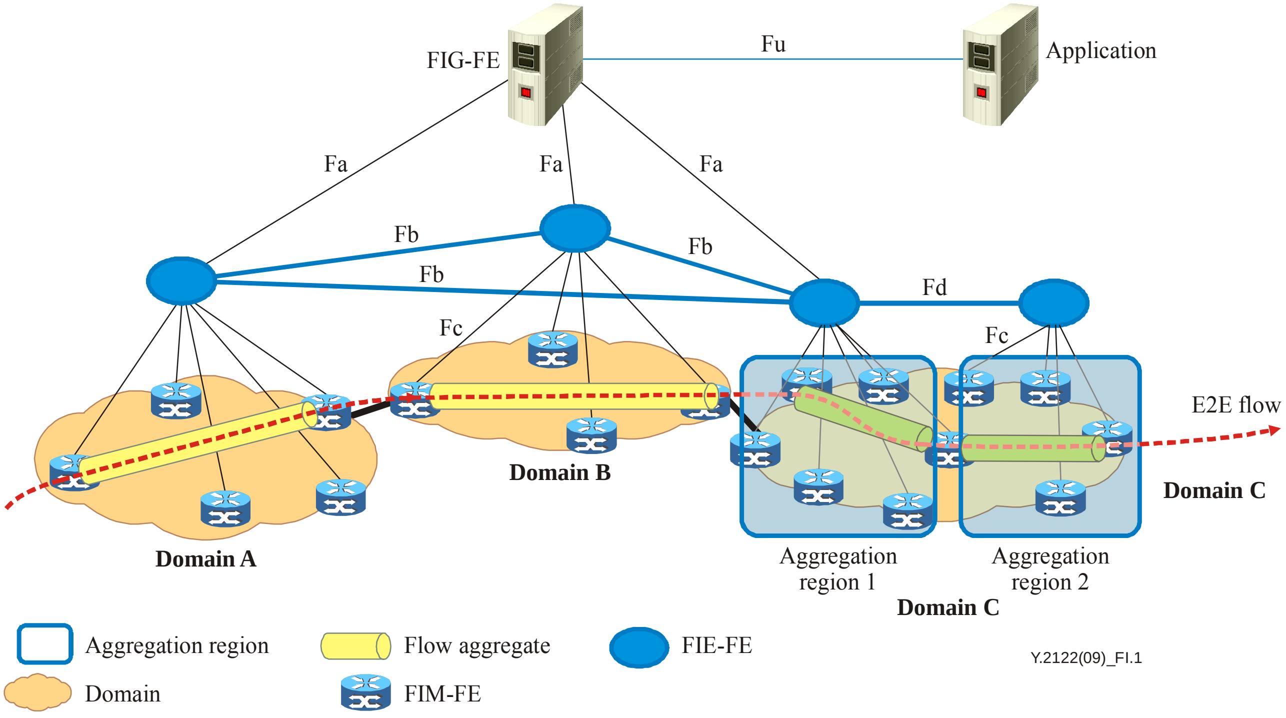

Figure I.1 – Example of implementation topology for FA information exchange

Appendix II

A usage example – Admission decision process

(This appendix does not form an integral part of this Recommendation)

The network parameter exchange can be used in the admission decision process: Should a newly requested flow be accepted? Given the end-to-end performance budget, such as the IPTD value for Class 0 traffic: 100 ms, one can decide to accept the flow if the networks N_1, N_2, … N_n in the paths can guarantee d_l ms, d_2 ms, …, d_n ms mean delays, and these sum up into a value less than 100 ms. As it has been pointed out in the Introduction, actual mean delay for this flow at each network cannot be measured prior to acceptance. The estimation or the prediction procedure should be conducted. Either of the two methods can be used for this purpose:

- Collaborative estimation method:
  1) Network parameters for flow aggregates are exchanged between neighbouring networks.
  2) A flow (f) requests admission, and during the call set-up it is determined to which flow aggregates (F_1, F_2, …, F_n) the flow will belong in the networks in the path.
  3) Based on the exchanged information on F_1, F_2, …, F_n, the delays d_1, d_2, …, d_n of F_1, F_2, …, F_n are estimated at the networks N_1, N_2, …, N_n, respectively. The estimation can be divided into two processes: the calculation based on the parameter values, and the adjustment based on the past monitoring.
  4) N_i gathers information of d_1, d_2, …, d_n. So does any other network in the path.
  5) N_i can decide whether the flow can meet the requirements. So can any other network in the path.
- Individual estimation method:
  1) Network parameters for flow aggregates are exchanged between neighbouring networks.
  2) A flow (f) requests admission, and during the call set-up it is determined to which flow aggregates (F_1, F_2, …, F_n) the flow will belong in the networks in the path.
  3) Based on the exchanged information on F_1, F_2, …, F_n, the delays d_1, d_2, …, d_n of F_1, F_2, …, F_n are estimated at network N_i, and also at N_{i+1} independently of N_i, and so on. The estimation can be divided into two processes: the calculation based on the parameter values, and the adjustment based on the past monitoring.
  4) N_i can decide whether the flow can meet the requirements. So can any other network in the path.

Note that the collaborative method requires a uniform framework for estimating the mean delay for every network in the path. Otherwise, even in exactly the same situation, different networks may generate a different estimated delay. Generally, it is assumed that the estimation processes are different from network to network, therefore the individual estimation method is preferred.

The following is the estimation example for d_i, for different types of networks, using the individual estimation method. Throughout this appendix the deterministic delay bound analysis technique has been adopted. Other analysis methods such as statistical delay analysis can also be used.

Note that the administrators can adjust later the worst case delay bound into a more realistic one, for example, by observing the real delay distribution after the call has been accepted. By doing so, a mapping rule from the D to the mean delay can be established over time.

II.1 DiffServ expedited forwarding

The delay bounds in a general topology network with the expedited forwarding PHB of the DiffServ architecture can be found in the bibliography. Let us denote ρ_i and σ_i as the leaky bucket parameters: the mean data rate and the maximum burst size of flow *i*, respectively. Assume that all the flows with premium service are constrained with leaky bucket, and also assume that the premium service traffic receives strict scheduling priority over other traffic. An exemplary delay (*D*) bound obtained in [b-Charny 00], [b-Wang 04], under the condition that $\alpha < 1/(H - 1)$, and under assumption that the incoming links' capacity is infinite, and given as:

$$D \leq \frac{H}{1-(H-1)\alpha}(\Delta + \tau) \qquad \text{(II-1)}$$

where *H* is the maximum hop count in the network, $\Delta = \max_S(L/r^S)$, r^S is the capacity of server *S*, and *L* is the maximum packet size. τ and α are defined by the inequalities $\sum_{j\in F_S} \sigma_j \leq \tau r^S$, and $\sum_{j\in F_S} \rho_j \leq \alpha r^S$, for any server *S* in the network with capacity r^S, in which there is a set of flows, F_S. One may call τ the burst allowance level measured in time for their transmission, and α the network utilization.

For a DiffServ network, the above equation for the delay bound of packets within the highest priority traffic can be used for inferring the IPTD. For example, given the hop count of a flow within a network (*H*), the network utilization (α), and the burst allowance level (τ), the delay characteristics can be inferred.

The key difficulty lies in calculating τ. In order to obtain this value, one has to know the maximum burst size (σ_i) at the entrance of the network of every flow sharing the output queue, or the server, in a network. Other parameters such as α, Δ, and *H* can be obtained from the traffic descriptor [b-ITU-T Y.1221] or from the inherited network characteristics. The σ_i at the entrance of the network (burst-in) should therefore be provided from the neighbouring networks.

II.2 IntServ guaranteed rate service

IntServ is now an obsolete QoS architecture which is not used in practice, except for its signalling protocol, RSVP. The myriad performance analyses on the end-to-end QoS are based mostly on IntServ, particularly on guaranteed service. The ideas from IntServ, such as resource reservation, flow protection, ingress regulation, etc. have been the key components of the numerous basic QoS provisioning architectures. It is, therefore, worth looking into the performance of IntServ networks, especially the delay bound.

The rate-guaranteeing schedulers (or servers), formerly called guaranteed-rate (GR) servers [b-LeBoudec 01] or latency-rate (LR) servers [b-Stiliadis 98] play the key role in bounding the end-to-end delay. The end-to-end delay of a flow *i* in a network with LR servers can be obtained by the following inequality from [b-Stiliadis 98]:

$$D_i \leq \frac{\sigma_i - L_i}{\rho_i} + \sum_{j=1}^{H} \Theta_i^{Sj} \qquad \text{(II-2)}$$

where Θ_i^{Sj} is the latency of the j_{th} server (S_j) for the flow *i* along the path of *i* in the network. The latency is the parameter of an LR server that is uniquely given. For example, the well-known packetized generalized processor sharing (PGPS) server or its implementation equivalent weighted fair queuing (WFQ) server has the latency:

$$\Theta_i^{PGPS} = \frac{L_i}{\rho_i} + \frac{L_{\max}}{r} \qquad \text{(II-3)}$$

where L_i is the maximum packet length of flow i, and L_{max} is the maximum packet length of all the flows in the server. Another popular scheduler of LR server class is the Deficit Round Robin (DRR). Its latency is given as follows.

$$\Theta_i^{DRR} = \frac{3F - 2\phi_i}{r} \tag{II-4}$$

where ϕ_i is the quantum value, or the relative portion of the server capacity given to flow i, and F is the sum of the ϕ_j for all the flows within the server.

As can be seen from equations II-2, II-3, and II-4, the end-to-end delay bound of an IntServ network can be calculated only from the parameters in the traffic descriptor, except σ_i, the maximum burst size of flow i at the entrance of the network, or the burst-in, which is different from what the flow has described at the UNI.

The "Pay burst only once" principle, however, comes in here. If two or more contiguous networks implement the IntServ-type rate-guaranteeing schedulers for a flow, the flow is protected throughout the networks, and these networks can be thought of as a single network and the delay component due to σ_i in equation II-2 is applied only once to the delay bound calculation. This principle is called "Pay burst only once".

II.3 Networks with various granularity flow aggregation (MPLS, FSA in [b-ITU-T Y.2121], etc.)

Theoretically, the rate-guaranteeing servers in the IntServ architecture will give exactly the same amount of service to a flow aggregate and to a micro-flow, given that the allocated rates are the same. Therefore, within a single aggregation region, the delay bound of a flow aggregate i can be calculated from equation II-2 as well. In this case, the maximum burst size σ_i is the sum of all the maximum burst sizes of the flows within the aggregate, and the mean data rate ρ_i is, again, the sum of all the mean data rates of the flows within the aggregate.

It is noteworthy that for a flow aggregate, when compared to a flow within the end-to-end delay bound can be reduced.

II.4 Calculation of the burst-in or equivalently the burst-out

As it has been observed, one of the most important parameters that has to be exchanged to calculate the delay bound of a network is the burst-in value of flows.

Therefore, a network has to be able to provide the burst-out value of a flow to the neighbouring networks. It has been suggested [b-Stiliadis 98] that the burst-out of a network can be calculated by:

$$\sigma_i^{out} = \sigma_i^{in} + \rho_i \sum_{k=1}^{H} \Theta_i^{S_k} \tag{II-5}$$

where σ_i^{in} and σ_i^{out} are the burst-in and burst-out of the flow i, respectively, and $\Theta_i^{S_k}$ is the latency of the k_{th} server (S_k) for the flow i along the path of i in the network. The latencies of servers in different QoS architectures are given as follows.

For a DiffServ network, the latency for a flow i in a strict priority queuing server S is given as:

$$\Theta_i^S = \frac{\sigma^S - \sigma_i^S + 2L}{r^S} \tag{II-6}$$

where σ^s is the sum of all the maximum burst sizes of the flows that share the server with i. It is now clear that the burst-out of a flow is a function of the burst-ins of the flows that the flow under observation (i) has met in the path in a network.

For a DiffServ network, therefore, it is required information from neighbouring networks such as the burst-in of the incoming flows, other than parameters specified in the traffic descriptor. The burst-ins of flows, or similarly the burst-outs of flows into another network, however, can be burdensome to calculate for a core network. There can be an approximation approach for calculating burst-ins from other networks.

For an IntServ network, the latencies of various servers are given, and two of them are listed in equations II-3 and II-4.

For networks with flow aggregation, under conditions that all the flows in a flow aggregate are leaky bucket constrained in front of an LR server *S*, and the aggregated data rate is less than the link capacity, the LR server for the flow aggregate is still an LR server for individual flows, with latency given as follows:

$$\Theta_i^S = \frac{\sigma_I^S - \sigma_i^S}{\rho_I} + \Theta_I^S \qquad \text{(II-7)}$$

where *I* is the flow aggregate, to which *i* belongs in the server.

Appendix III

A usage example – Flow granularity and continuity decision process

(This appendix does not form an integral part of this Recommendation)

For networks having flow aggregation capability with variable granularity and continuity, QoS consideration should be an issue when it comes to the decision on the granularity and continuity of a flow aggregate. The reason for this is that QoS performance depends heavily upon these characteristics of a flow aggregate. It is generally well understood that the FA with large granularity lacks protection to an individual flow from other bursty traffic. Smaller scale FA is therefore preferable, yet it has to overcome the scalability problem. It is again generally understood that the FA with longer continuity gives a better delay performance, yet with less scalability. The network operator has to decide about the ubiquitous performance-scalability compromise. How many IP flows should be in an LSP? How long is the path of an LSP? One can simply put all the IP flows with the same source-destination pair into a single LSP. One can go further to put traffic type into consideration, therefore IP flows with the same source-destination pair and the same application may only be in a same FA. On the other hand, when the network is large, one can avoid having too many FAs to handle by dividing a network into several subnetworks, therefore reducing continuity and increasing granularity.

In terms of QoS performance, it is sometimes enough for a network not to violate an allocated performance margin for a flow. In other words, if a network knows that the sum of impairments incurred by the other networks in the path is safely lower than what is requested, the network may shorten the continuity of the FA, given that all the flows' requests are sufficiently met. The following FA granularity and continuity decision process may be adopted for efficient complexity management for a network.

III.1 FA granularity and continuity decision process

- Network parameters for flow aggregates are exchanged between neighbouring networks.
- Based on the exchanged information, for a flow in FA (F_i), the performance measures (e.g., delay) in the path (p^i_1, p^i_2, …, pin) are estimated. The estimation can be divided into two processes: the calculation based on the parameter values, and the adjustment based on the past monitoring. Similarly, for all i in the FA, the performance measures are estimated.
- For the FA (F), it has been discovered that all the flows inside F enjoy sufficient margins in performance requirements, throughout the paths; or for the FA (F), if its performance per hop in the current network is sufficiently different from those of the neighbouring networks, the network can decide whether the FA can alter its granularity and continuity.

Appendix IV

FIM-FE usage scenarios

(This appendix does not form an integral part of this Recommendation)

In this appendix, some of the scenarios regarding FIM-FE usage are considered.

IV.1 FA table entry writing

The proper parameters for a given FA can be obtained by either a signalling or a management procedure. For example, if FIM-FE is at the ASP, then ASP can send a request message to the downstream nodes in the AR. The nodes within the AR in response fill the tables. The AEP can return a reply message back to the ASP.

On the contrary, FA parameters may not be obtained from individual nodes. Rather, it can be induced from the network's flow aggregation policy. For example, if a network aggregates incoming flows using a strict rule, such that the sum of sustainable rates does not exceed a pre-defined ratio of the total capacity, this rule itself can be written in the corresponding field in the FA table as a maximum value. With such an induction, the accuracy of the information written in a FA table will be somewhat downgraded. The obvious gain from the loss of accuracy is more scalability.

IV.2 Information management mechanism for multiple ARs in a single network

In case where there are multiple ARs for a flow in a network, the FIM-FEs are required to reside per AR, as in Figure IV.1.

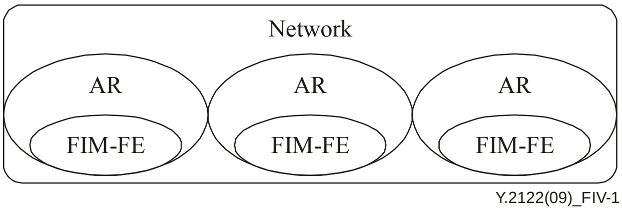

Figure IV.1 – FIM-FEs for multiple ARs in a network

IV.3 Information management mechanism for a single AR over multiple networks

Similarly, there can be cases where a single AR covers multiple network domains, for example with inter-domain MPLS LSPs [b-IETF RFC 4216]. In such a case an FIM-FE is required to reside only for an AR, not for a network, as depicted in Figure IV.2.

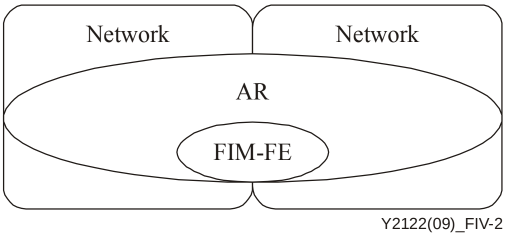

Figure IV.2 – FIM-FEs for single AR in multiple networks

Appendix V

FIE-FE usage scenarios

(This appendix does not form an integral part of this Recommendation)

In this appendix, some of the scenarios regarding FIE-FE usage are considered.

V.1 Cases of information exchange for multiple-ARs in a network

Since a requesting entity does not have information on the identity of a FA on a requested network, the request query is likely to be based on flow identity. The requested network is required to respond with the corresponding FA table(s).

If there are multiple ARs in a network for a flow (a good example of such multiple ARs in a network would be one with MPLS LSP aggregation [b-IETF RFC 4206] capability), the corresponding FA tables for a specific flow are recommended to be handled by a single FIE-FE per network. The FIE-FE responsible for a specific FA can be either within an aggregation region (AR) or in a central management entity in the network. If the FIE-FE is within an AR, it can be either at the aggregation start point (ASP) or at the aggregation end point (AEP).

V.2 Cases of information exchange for a single AR in multiple networks

If the requested entity does not have an FIM-FE within the network, because the FA's AR covers multiple networks and the corresponding FIM-FE is on another network, then the requested network is required to be able to get the FA information from the FIM-FE on another network. Therefore the Fd interface can be formed over network boundaries, as depicted in Figure V.1.

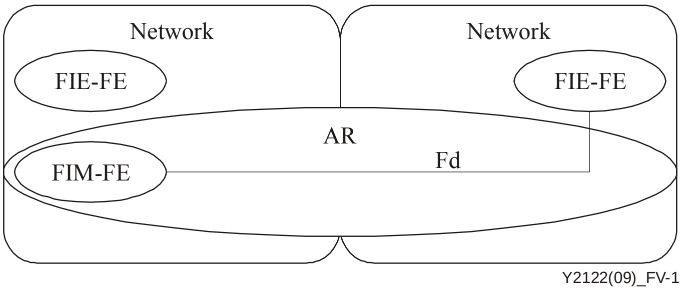

Figure V.1 – Fd interface over network boundary

Appendix VI

Example allocation of FA parameter values

(This appendix does not form an integral part of this Recommendation)

In this appendix, an example strategy for allocating FA parameter bits and values is recommended.

VI.1 Mapping rule from flow ID to FA ID

This field is recommended to have 8 bits. The following table shows the assigned binary code with the MSB four bits to the transport technology of the network and the mapping policy.

Table VI.1 – Recommended bits assignment for mapping rule

	MSB (Bit 7)	**Bit 6**	**Bit 5**	**Bit 4**
Best-effort	0	0	0	0
DiffServ	0	1	0	X
DiffServ with TE	0	1	1	X
MPLS and its variations	1	0	X	X
FSA	1	1	X	X
Reserved	0	0	1	X

The LSB four bits will indicate technology-specific mapping rule. The detailed assignment is for further study.

VI.2 Bandwidth allocation enforcement method or scheduling method

This field is recommended to have 8 bits.

Table VI.2 – Recommended bits assignment for scheduling method

	MSB (Bit 7)	**Bit 6**	**Bit 5**
Rate-guaranteeing schedulers	0	X	X
Priority schedulers	1	0	0
Class based bandwidth sharing schedulers (DiffServ-TE variations)	1	0	1
Reserved (including non work-conserving schedulers)	1	1	X

The schedulers will be further specified by the remaining five bits. This further categorization is for future work.

VI.3 Representative (R)

This field is recommended to be 1 bit.

VI.4 FA identifier

This field is recommended to have 31 bits.

VI.5 Number of flows within the FA

This field is recommended to have 8 bits. This field is recommended to strictly follow the normalized exponential notation. The most significant four bits represent the coefficient, in decimal, and the least significant four bits represent the exponent, also in decimal. Therefore, for example, if there are 2340 flows within a FA, then this field is recommended to be "0010_0011" which represents 2 x 10^3. Similarly, "1111_0100" represents 1.5 x 10^8 flows within a FA.

Note that the most significant four bits "1011" represent the coefficient 1.1, not 11. If the actual coefficient is more than or equal to 1.6 and less than 2.0, then the most significant four bits are suggested to be 0010, i.e., 2 in decimal. The range of this field can be from 0 to 9 x 10^{15} in decimal.

The value "0000_0000" means that the real value of this field is unknown.

VI.6 Maximum packet length divided by link capacity

This field is recommended to have 8 bits. The unit of this field is "seconds". This field is recommended to follow the normalized exponential notation. The most significant four bits represent the coefficient, in decimal, and the least significant four bits represent the exponent, also in decimal. The exponent is assumed to be a negative number without any extra bit to mark. Therefore, for example, 1.2 x 10^{-5} can be written as "1100_0101".

The range of this field can be from 1 x 10^{-15} to 9 in decimal.

The value "0000_0000" means that the real value of this field is unknown.

VI.7 Number of hops within the aggregation region or the policy region

This field is recommended to have 8 bits. The range of this field can be from 0 to 255.

The value "0000_0000" means that the real value of this field is unknown.

VI.8 Maximum number of aggregation regions

This field is recommended to have 8 bits. The range of this field can be from 0 to 255. The value "0000_0000" means that the real value of this field is unknown.

VI.9 Maximum sum of sustainable transfer rates (Rs) divided by link capacity

This field is recommended to have 8 bits. This field specifies the ratio of the link capacity to the nominal sustainable rate of the FA. It is unlikely for the value of this field to exceed one. This field is recommended to follow the normalized exponential notation. The most significant four bits represent the coefficient in decimal and the least significant four bits represent the exponent also in decimal. The exponent is assumed to be a negative number without any extra bit to mark. Therefore, for example, 1.2 x 10^{-5} can be written as "1100_0101".

The range of this field can be from 1 x 10^{-15} to 9 in decimal.

The value "0000_0000" means that the real value of this field is unknown.

VI.10 Maximum sum of sustainable burst sizes (Bs) divided by link capacity

This field is recommended to have 8 bits. This field specifies the ratio of the link capacity to the nominal sustainable burst size of the FA. This field is recommended to follow the normalized exponential notation with sign. The most significant four bits represent the coefficient, in decimal, the least significant bit represents the sign of the exponent, and the remaining three bits represent the exponent, also in decimal. Therefore, for example, 1.2 x 10^{-5} can be written as "1100_101_1". The least significant bit 1 represents a minus sign.

The range of this field can be from 1 x 10^{-7} to 9 x 10^7 in decimal.

The value "0000_0000" means that the real value of this field is unknown.

VI.11 Minimum guaranteed service rate divided by sum of Rs (if rate-based) or scheduling priority

This field is recommended to follow the normalized exponential notation. The most significant four bits represent the coefficient and the least significant four bits represent the exponent of the minimum guaranteed service rate divided by the sum of Rs. The exponent is assumed to be always negative or zero. Therefore, for example, 1.2 x 10^{-5} can be written as "1100_0101".

The range of this field can be from 9 to 1 x 10^{-15} in decimal.

The value "0000_0000" means that the real value of this field is unknown.

VI.12 Packet discard priority

This field is recommended to have 4 bits. The range of this value is 0 to 255.

VI.13 Class (service context)

This field is recommended to have 8 bits. The range of this value is 0 to 255.

VI.14 Summary

Table VI.3 summarizes the example parameter bits allocated in this appendix.

Table VI.3 – Recommended bit assignment for parameters

Taxonomy	Field	Number of bits	Range
Parameters not specific to a FA	Mapping rule (policy) from Flow ID (IP 5-tuple) to FA ID	8	
	Bandwidth allocation enforcement method or (scheduling method)	8	
FA parameters	Representative (R)	1	0 to 1
	FA Identifier	31	0 to 2^{31}–1
	Number of flows within the FA	8	0 to 9×10^{15}
	Maximum packet length divided by link capacity	8	1×10^{-15} to 9
	Number of hops within the aggregation region or the policy region	8	0 to 255
	Maximum number of aggregation regions	8	0 to 255
	Maximum sum of sustainable transfer rates (Rs) divided by link capacity	8	1×10^{-15} to 9
	Maximum sum of sustainable burst sizes (Bs) divided by link capacity	8	1×10^{-7} to 9×10^{7}
	Minimum guaranteed service rate divided by sum of Rs (if rate-based) or scheduling priority	8	9 to 1×10^{-15}
	Packet discard priority	8	0 to 255
	Class (service context)	8	0 to 255

Appendix VII

Network specific parameter values

(This appendix does not form an integral part of this Recommendation)

The purpose of this appendix is to recommend how the values of parameters are determined for each transport technology. The parameters defined in clause 8 can be interpreted differently depending on the network. How these fields can be interpreted in different networks will be elaborated in separate subclauses.

The examples of the DiffServ, DS-MPLS [b-IETF RFC 3270] networks, FSA networks are considered. It is also considered how the FA parameters are assigned with the inherent network parameters within such networks.

VII.1 DiffServ network specific parameters

VII.1.1 Representative (R)

In DiffServ networks, an AR contains usually only a single node. As a consequence, the management of the FA table can be cumbersome, although the number of FAs in a node is far smaller compared to other types of networks. It is expected that a DiffServ network may easily set this field to true.

VII.1.2 FA identifier

If the R field is false, this field in a DiffServ network is recommended to indicate 6-bit DiffServ code point (DSCP) [b-IETF RFC 2474]. The matching FA identifier will be the six bits from bit 7 to bit 2. Optionally, the two explicit congestion notification (ECN) bits may be copied to bit 1 and bit 0 of this field.

VII.1.3 Number of flows within the FA

If the R field is false, this field in a DiffServ network is recommended to indicate the number of flows within the behaviour aggregate with the same DSCP.

VII.1.4 Packet discard priority

This field corresponds to the drop precedence field in DSCP. It effectively indicates the buffer space allocation priority among FAs. Since the DSCP is carried in the FA identifier field, this field may be set to zero, i.e., "0000_0000".

VII.1.5 Class (service context)

Since the DSCP is carried in the FA identifier field, this field may be set to zero, i.e., "0000_0000".

VII.1.6 Other fields

If the R field is false, the other fields are recommended to represent corresponding information, considering the BA as the FA.

VII.2 DiffServ with traffic engineering (TE) network specific parameters

VII.2.1 FA identifier

The FA identifier in a DiffServ-TE network is recommended to represent, among the 32-bit top level (if hierarchical header is being used) label stack; a 20-bit label value, a 3-bit field for QoS (quality of service) priority (experimental), and a 1-bit bottom of stack flag. The matching FA identifier is suggested to be from bit 23 to bit 0.

VII.2.2 Minimum guaranteed service rate divided by sum of Rs (if rate-based) or scheduling priority

The right value of this field for DiffServ-TE networks is intuitively unclear, due to the sharing nature of the scheduling policies adopted in these networks. Current dominant practices of bandwidth sharing methods include the Russian Dolls model (RDM) [b-IETF RFC 4127] and the maximum allocation model (MAM) [b-IETF RFC 4125]. Other mechanisms are being proposed as well (e.g., maximum allocation with reservation (MAR) [b-IETF RFC 4126]).

This field in DiffServ-TE networks is recommended to indicate "the average minimum guaranteed bandwidth given to the LSP (or FA)" by the following calculation.

For both RDM and MAM, this field is recommended to indicate the 'Reserved(CTi)' value divided by the number of LSPs from CTi, where CTi is the Class Type to which the current LSP (or FA) belongs. The eight bits will indicate this value with normalized exponential notation. 'Reserved(CTi)' is defined as the total amount of the bandwidth reserved by all the established LSPs which belong to CTi.

In RDM, for the Class Type i (CTi), with i from 0 to 7, the total amount of bandwidth reserved from CT7 to CTi (Reserved (CT7) + … + Reserved (CTi)) must be less than BCi. BC0 is then defined as the maximum reservable bandwidth. Therefore, all LSPs from CT7 use no more than BC7; all LSPs from CT6 and CT7 use no more than BC6; all LSPs from CT5, CT6 and CT7 use no more than BC5; etc. In MAM, the sharing mechanism is a bit simpler such that the all LSPs from CTi use no more than BCi. The key sharing idea here is that the sum of Reserved(CTi) over all i must be less than or equal to the maximum reservable bandwidth. For a network management entity, specifying a real value of the assigned bandwidth for a single LSP is hard, if not impossible, because it dynamically changes over time. Therefore, in both practices (RDM and MAM), specifying a worst case is the best alternative, even if it may be far from the bandwidth actually assigned to a single LSP.

VII.2.3 Class (service context)

The Class Type of the LSP (or FA) is recommended to be specified in this field. Each DSCP value for BA indication in DiffServ (e.g., Decimal 10 for AF11 BA) should be used intact. The matching bits to DSCP are recommended to be the six bits from bit 7 to bit 2. Optionally, the two explicit congestion notification (ECN) bits may be copied to bit 1 and bit 0 of this field.

VII.2.4 Other fields

If the R field is false, the other fields are recommended to represent corresponding information, considering the LSP (E-LSP or L-LSP) as the FA.

VII.3 Flow state aware (FSA) network specific parameters

VII.3.1 FA identifier

In FSA networks, ingress edge nodes may support the option to aggregate selected flows into a flow aggregate, based on some criteria such as the service context, the preference indicator value, and the path in the network. The identifier for the flow aggregate, however, is not specified in FSA networks. Instead it is recommended to use an MPLS label for the identifier. In such a case, the FA identifier for FSA networks is suggested to be the MPLS label as well.

The FA identifier in FSA networks is recommended to represent, among the 32-bit top level (if hierarchical header is being used) label stack; a 20-bit label value, a 3-bit field for QoS (quality of service) priority (experimental), and a 1-bit bottom of stack flag. The matching FA identifier is recommended to be from bit 23 to bit 0.

VII.3.2 Packet discard priority

This field is recommended to represent the packet discard priority (with the same name in this Recommendation) field defined in FSA networks. Since the exact usage of header fields for FSA networks must be specified, it is to be decided how this field will be specified as well.

VII.3.3 Class (service context)

This field is recommended to represent the service context (with the same name in this Recommendation) field defined in FSA networks. Since the exact usage of header fields for FSA networks must be specified, it is to be decided how this field will be specified as well.

VII.3.4 Other fields

If the R field is false, the other fields are recommended to represent corresponding information.

Bibliography

[b-ITU-T Y.1221] Recommendation ITU-T Y.1221 (2002), *Traffic control and congestion control in IP-based networks*.

[b-ITU-T Y.2121] Recommendation ITU-T Y.2121 (2008), *Requirements for the support of flow-state-aware transport technology in NGN.*

[b-ITU-T Y.2173] Recommendation ITU-T Y.2173 (2008), *Management of performance measurement for NGN.*

[b-IETF RFC 2246] IETF RFC 2246 (1999), *The TLS Protocol Version 1.0*.

[b-IETF RFC 2474] IETF RFC 2474 (1998), *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*.

[b-IETF RFC 2475] IETF RFC 2475 (1998), *An architecture for Differentiated Services*.

[b-IETF RFC 3270] IETF RFC 3270 (2002), *Multi-Protocol Label Switching (MPLS) Support of Differentiated Services*.

[b-IETF RFC 4125] IETF RFC 4125 (2005), *Maximum Allocation Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering*.

[b-IETF RFC 4126] IETF RFC 4126 (2005), *Max Allocation with Reservation Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering & Performance Comparisons*.

[b-IETF RFC 4127] IETF RFC 4127 (2005), *Russian Dolls Bandwidth Constraints Model for Diffserv-aware MPLS Traffic Engineering.*

[b-IETF RFC 4206] IETF RFC 4206 (2005), *Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)*.

[b-IETF RFC 4216] IETF RFC 4216 (2005), *MPLS Inter-Autonomous System (AS) Traffic Engineering (TE) Requirements*.

[b-IETF RFC 4301] IETF RFC 4301 (2005), *Security Architecture for the Internet Protocol.*

[b-IETF RFC 5470] IETF RFC 5470 (2009), *Architecture for IP Flow Information Export.*

[b-Charny 00] A. Charny and J.-Y. Le Boudec (2000), *Delay bounds in a network with aggregate scheduling* – First International Workshop of Quality of Future Internet Services (QOFIS2000).

[b-Le Boudec 01] J.-Y. Le Boudec and P. Thiran, *"Network Calculus – A Theory of Deterministic Queueing Systems for the Internet," Springer, LNCS vol. 2050, 2001*.

[b-Stiliadis 98] D. Stiliadis and A. Varma, *"Latency-Rate servers: A general model for analysis of traffic scheduling algorithms", IEEE/ACM Trans. Networking, vol. 6, no. 5 Oct. 1998*.

[b-Wang 04] Shengquan Wang; Dong Xuan; Bettati, R.; Wei Zhao, *Providing absolute differentiated services for real-time applications in static-priority scheduling networks,* IEEE/ACM Transactions on Networking, Volume 12, Issue 2, April 2004 Page(s): 326 – 339.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Terminals and subjective and objective assessment methods
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	**Global information infrastructure, Internet protocol aspects and next-generation networks**
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2010