International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Y.3042

(04/2013)

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

Future networks

Smart ubiquitous networks – Smart traffic control and resource management functions

Recommendation ITU-T Y.3042

International Telecommunication Union

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
Packet-based Networks	Y.2600–Y.2699
Security	Y.2700–Y.2799
Generalized mobility	Y.2800–Y.2899
Carrier grade open environment	Y.2900–Y.2999
FUTURE NETWORKS	**Y.3000–Y.3499**
CLOUD COMPUTING	Y.3500–Y.3999

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

Recommendation ITU-T Y.3042

Smart ubiquitous networks – Smart traffic control and resource management functions

Summary

Recommendation ITU-T Y.3042 specifies the smart traffic control and resource management functions for SUN. It defines the motivation and identifies the high-level requirements and functional architecture for providing relevant network capabilities. This Recommendation also identifies mechanisms in terms of "smart and ubiquitous" aspects of networks.

History

Edition	Recommendation	Approval	Study Group
1.0	ITU-T Y.3042	2013-04-13	13

Keywords

Smart resource management, smart traffic control, smart ubiquitous network, SUN.

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/.

Table of Contents

Page

Recommendation ITU-T Y.3042

Smart ubiquitous networks – Smart traffic control and resource management functions

1 Scope

This Recommendation specifies smart traffic control and resource management functions to provide fair usage of network resources using context awareness capability in smart ubiquitous networks (SUN). This Recommendation covers the following:

- Motivation and objectives of smart traffic control and resource management;
- Requirements for smart traffic control and resource management for SUN;
- High-level architecture and functional architecture;
- Control and management mechanisms.

NOTE 1 – Usage of context information is optional and the mechanism details of context awareness capability are outside of scope in this Recommendation.

NOTE 2 – Regulatory aspects of monitoring (e.g., depth of IP packet allowed to be monitored) are outside of scope in this Recommendation.

NOTE 3 – Traffic, means user or service provider traffic and resource, means operational status of network elements (e.g., CPU usage, interface utilization, etc.) in this Recommendation.

NOTE 4 – Controlling traffic includes both application and network layer control. Some examples of application layer controls are video pacing, video transcoding and/or HTTP adaptive streaming (HAS).

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 E.600] Recommendation ITU-T E.600 (1993), *Terms and definitions of traffic engineering.*

[ITU-T Q.9] Recommendation ITU-T Q.9 (1988), *Vocabulary of switching and signalling terms.*

[ITU-T Y.1541] Recommendation ITU-T Y.1541 (2011), *Network performance objectives for IP-based services.*

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

[ITU-T Y.3041] Recommendation ITU-T Y.3041 (2013), *Smart ubiquitous networks – Overview.*

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 smart ubiquitous networks (SUN) [ITU-T Y.3041]: IP-based packet networks that can provide transport and delivery of a wide range of existing and emerging services to people and things. The services provided by the SUN can cover aspects such as control, processing and storage.

NOTE 1 – The network is smart in the sense that it is knowledgeable, context-aware, adaptable, autonomous, programmable, and can perform services effectively and securely.

NOTE 2 – The network is ubiquitous in the sense that it allows access anytime, anywhere, through varied access technologies, access devices, including end user devices, and human-machine interfaces.

3.1.2 context [ITU-T Y.3041]: The information that can be used to characterize the environment of a user.

3.1.3 busy hour [ITU-T E.600]: The continuous 1-hour period lying wholly in the time interval concerned for which the traffic or the number of call attempts is greatest.

3.1.4 (signalling) traffic flow control [ITU-T Q.9]: Actions and procedures intended to limit signalling traffic at its source in the case when the signalling network is not capable of transferring all signalling traffic offered by the User Parts due to network failures or overload situations.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 data cap: Upper limit of the total traffic volume allowed by the SLA contracted between the traffic source (e.g., ISP) and destination (e.g., user).

3.2.2 fair usage: Equal treatment for the same service(s), including application(s), between different users (e.g., end-user, applications) with the same SLA.

3.2.3 heavy service traffic: Traffic level measured in the volume of specific service(s) which overloads network resources at the interface from/to service providers. It has an impact on other service traffic on the same network resource (e.g., degrading the quality).

3.2.4 heavy signalling traffic: Traffic level measured in the volume and frequency that triggers signalling traffic flow control [ITU-T Q.9] (e.g., heavy traffic level caused by keep-alive messages).

3.2.5 heavy user traffic: Traffic level measured in the volume of traffic initiated by an end-user who overloads network resources at the interface from/to the end-user. It affects the traffic quality of other users.

3.2.6 surge traffic: Traffic level measured in the aggregated volume initiated by multiple users which overloads network resources at special events such as an abrupt occurrence of bad weather, a particular sports or music event that attracts a high number of users, etc.

3.2.7 monopolization: Condition of network resource overloads caused by heavy service or user traffic. It affects the traffic quality of other service(s) or user(s).

3.2.8 over-sized traffic: Traffic level that is larger in volume than the recommended size for a service. The recommended volume of such a service should be identified by SLA between the traffic source and destination. It is a special case of heavy service traffic.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

3DTV	3 Dimensional TV
CP	Contents Provider
DSCP	Differentiated Services Code Point
HDTV	High Definition TV
IoT	Internet of Things
IPTV	Internet Protocol TV
ISP	Internet Service Provider
OTT	Over the Top
PDL	Progressive Down Load
POP	Point of Presence
QCI	QoS Class Identifier
QoE	Quality of Experience
QoS	Quality of Service
RA-FE	Resource Analysis Functional Entity
RMAF	Resource Monitoring and Analysis Function
RM-FE	Resource Monitoring Functional Entity
SCD-FE	Smart Correlation and Decision Functional Entity
SLA	Service Level Agreement
SRC-FE	Smart Resource Control Functional Entity
STC-FE	Smart Traffic Control Functional Entity
STCRMF	Smart Traffic Control and Resource Management Functions
STRCF	Smart Traffic and Resource Control Function
SUN	Smart Ubiquitous Network
TA-FE	Traffic Analysis Functional Entity
TMAF	Traffic Monitoring and Analysis Function
TM-FE	Traffic Monitoring Functional Entity
TRAF	Traffic Control and Analysis Function
TRCMF	Traffic Resource Control and Management Function
VNO	Virtual Network Operator

5 Conventions

In this Recommendation:

The keywords "is required to" indicate a requirement which must be strictly followed and from which no deviation is permitted, if conformance to this Recommendation is to be claimed.

The keywords "is prohibited from" indicate a requirement which must be strictly followed and from which no deviation is permitted, if conformance to this Recommendation is to be claimed.

The keywords "is recommended" indicate a requirement which is recommended but which is not absolutely required. Thus, this requirement need not be present to claim conformance.

The keywords "is not recommended" indicate a requirement which is not recommended but which is not specifically prohibited. Thus, conformance with this Recommendation can still be claimed even if this requirement is present.

The keywords "can optionally" indicate an optional requirement which is permissible, without implying any sense of being recommended. This term is not intended to imply that the vendor's implementation must provide the option, and the feature can be optionally enabled by the network operator/service provider. Rather, it means the vendor may optionally provide the feature and still claim conformance with this Recommendation.

6 Introduction

6.1 Motivation

Developments of various smart devices (e.g., smartphones, pads, IoT devices, etc.) in user environments of IP networks combined with newly emerging services such as Smart TV, 3DTV, and network gaming have led to increases in service traffic. These smart devices require networks to allocate more resources to support requested bandwidth and various service features including real-time, non-real-time and store-forward.

They are becoming one of the causes for "data explosion". In addition, a small number of users and service providers, generating heavy traffic, can monopolize most of the network's resources. Such monopolization needs to be controlled to ensure fair usage among users. This monopolization leads to degradation of quality of service and, in addition, discourages further advances in smart devices and related services development (see Appendix I). Therefore, more creative ways need to be identified to ensure fair usage of finite network resources.

In mobile environments, developments of smart user devices have enabled various broadband multimedia services which had previously been provided over the fixed network. Such changes in the mobile services cause network overload issues which can result in limited access resources for new users. In this environment, service traffic generated by service providers is transported over fixed networks and is delivered to the end-user devices over mobile access networks. Mobile access capacity is relatively limited compared to fixed access capacity. Therefore, it is noted that incoming traffic from fixed network environments to the mobile access environments needs to be managed for fair usage.

6.2 Objectives

Understanding the relationship among different factors, especially bandwidth, latency and operations complexity is essential for better operation of networks and services. Figure 6-1 shows three different areas of network operations based on the relationships of those factors. Their characteristics are as follows:

Area 1: Under-utilized case: The case where network resources are underutilized, QoS is easy to maintain and network management is simple. However, the network operator needs more users and high bandwidth consuming traffic for economic usage of network resources.

Area 2: Optimally utilized case: The case where optimal balance between high network resources utilization, QoS objectives and low network management complexity is maintained. It is noted that this is a recommended situation for an optimally operated network.

Area 3: Highly utilized case: case where network resources are highly utilized, QoS objectives may not be met and network management complexity is high. Many users in this area may experience difficulties in connecting to the networks and/or use of service. More especially, this situation gets worse in the case of mobile network environments. It is recommended to deploy traffic control and resource management for better service delivery and usage of network resources.

The objective of this Recommendation is to provide a traffic control and resource management function to address issues of Area 3, consistent with needs of Area 2. The needs of Area 1 are a business case and are outside of the scope of this Recommendation.

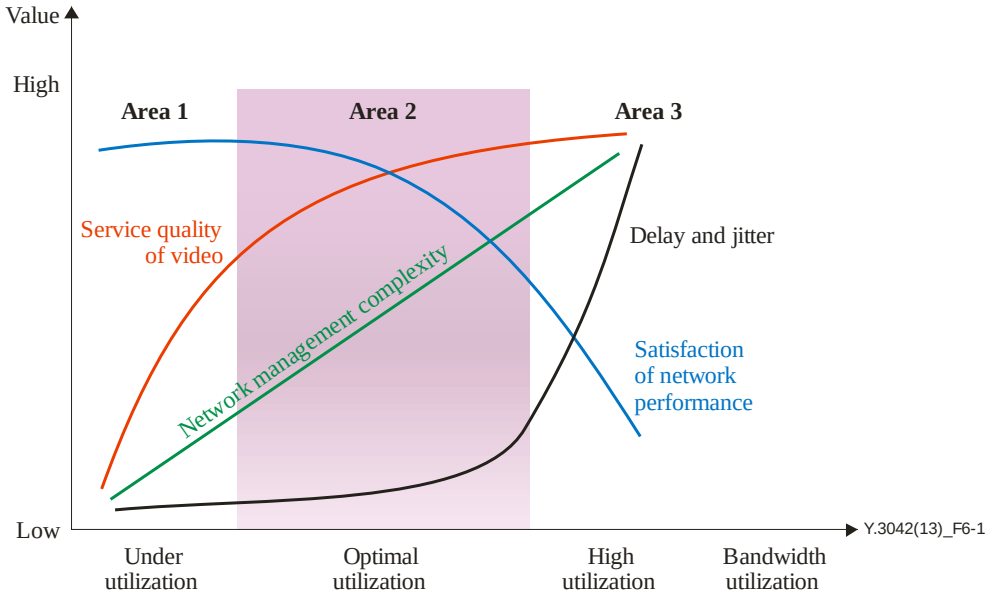

Figure 6-1 –Three different areas of network operation

7 Fine-grained classifications of traffic

There are many parameters used to specify traffic characteristics such as bandwidth, jitter, delay and burstiness, etc. They are defined in [ITU-T Y.1541]. SUN smart traffic control and resource management functions (STCRMF) use these parameters. In addition, [ITU-T Y.1541] defines new traffic parameters and classifications for newly emerged traffic types which have heavy traffic volumes and long duration characteristics generated by smart devices and service providers.

7.1 Fine-grained bandwidth

Under legacy telecommunication environments, traffic is classified into only two categories based on bandwidth – broadband and narrowband. Broadband refers to traffic above or equal to 2 Mbit/s and narrowband refers to those below 2 Mbit/s. This classification was widely accepted in the traditional telecommunication environments. In the SUN environment however, classification criteria have become more complicated due to the reasons described in clause 6.1.

Fine-grained bandwidth classification is defined to meet such needs. Three fine-grained bandwidth types in narrowband and two fine-grained bandwidth types in broadband are defined as the following bandwidth types:

– Type 0 (~ up to 1 kbit/s): It is used for very small amounts of data around several 100 bit/s levels. Traffic generated by sensors or signalling traffic such as heart-bit/beacon can belong to this type;

– Type 1 (1 ~ 128 kbit/s): It is used for traditional telecommunication services delivering voice and low quality images at less than 128 kbit/s. Some signalling traffic can belong to this type;

– Type 2 (128 kbit/s ~ 2 Mbit/s): It is used for typical multimedia services delivering video and high quality audio/voice/ image;

– Type 3 (2 ~ 20 Mbit/s): It is used for advanced multimedia services delivering HD and full HD quality video, etc.;

– Type 4 (above 20 Mbit/s): It is used for emerging services like 3D video services and advanced convergence services such as heart simulation and remote-medical surgery services.

7.2 Fine-grained service duration

Besides traffic volume, burstiness (e.g., burst or steady) and time sensitivity (e.g., real or non-real), another important factor which should be considered in SUN, is service duration. Since the introduction of always-on and fixed rate provision capabilities in certain services (e.g., IPTV, remote surgery, videoconferencing and network gaming), users are connected much longer regardless of actual usage of a service. In addition, ad-hoc applications such as P2P and online download allow users to be connected without any bandwidth concerns. Many of the above-mentioned services and applications have service duration over one hour with bandwidth types 3 and/or 4. Therefore, service duration should be considered as one of the important factors for traffic control and resource management.

Fine-grained service duration classification is defined to meet such needs. Five service duration types are defined in the range from very short period (e.g., less than one second) up to very long period (e.g., over one hour) as follows:

– Type 0 (less than one second): It is used for instantaneous data transmission (e.g., automatically generated sensor data transmission). Generally, this type of traffic does not have an impact on the operation of networks unless there is congestion caused by a sudden data surge (e.g., due to natural disaster);

– Type 1 (one second ~ less than ten minutes): It is used for traditional telecommunication services such as voice telephony, fax and messaging services. Most of them use bidirectional symmetrical communication services. It also includes unidirectional communication services such as messaging and music services;

– Type 2 (ten minutes ~ less than thirty minutes): It is used typically for web-based services. But it also includes voice and video telephony;

– Type 3 (30 minutes ~ less than one hour): It is used typically for video streaming service for which the contents are normal TV programmes (running time is less than one hour).It also includes voice and video conferencing;

– Type 4 (over one hour): It is used typically for video streaming service for which the contents are HD and/or 3D movies (running time is more than one hour). It also includes newly emerging services such as remote medical surgery, etc.

7.3 Fine-grained traffic classes

As identified in clauses 7.1 and 7.2, it is possible to map services in terms of "bandwidth" and "service duration". Fine-grained traffic classes are shown in Table 7-1. There are four traffic classes, class 0 to class 3. The details are described below. More specifically, classes 0, 2 and 3 are the particular focus for traffic control and resource management since they have significant impact on network performance.

Table 7-1 identifies fine-grained traffic classes:

– Class 0: It is characterized by bandwidth types 0 and 1 with service duration type 0. For example, some traffic of this class is data from sensors for e-health or for detecting natural disasters.

– Class 1: It is characterized by the concatenation of bandwidth types 0 to 4 with service duration type 1, bandwidth types 1 and 2 with service duration type 2, and bandwidth type 1 with service duration types 3 and 4. The majority of public services such as voice telephony belong to this traffic class.

– Class 2: It is characterized by concatenation of bandwidth types 3 and 4 with service duration type 2 and bandwidth types 2 to 4 with service duration types 3 and 4 except class 3. Many emerging services with advanced multimedia capabilities belong to this traffic class.

– Class 3: It is characterized by bandwidth type 4 with service duration type 4. High-bandwidth consuming services such as 3D remote medical surgery belong to this traffic class.

Table 7-1 – Fine-grained traffic classes

		Service duration				
		Type 0 (less than 1 second)	**Type 1** (1 second-less than 10 minutes)	**Type 2** (10 minutes-less than 30 minutes)	**Type 3** (30 minutes-less than 1 hour)	**Type 4** (over 1 hour)
Bandwidth	**Type 0** (up to 1 kbit/s)	Simple sensor data	Sensor data	NYI	NYI	NYI
	Type 1 (1 - 128 kbit/s)	Text (SMS) complicated sensor data	MMS Voice phone/ messaging	Voice phone	Voice phone Voice conference	Voice conference
	Type 2 (128 kbit/s - 2 Mbit/s)	NYI	Low-Q video messaging and video clip HQ music	Low-Q video phone/conferen. Inter-mediate size file transfer	P2P download Web TV Web casting	P2P download Web TV/casting Tele-Video-Surveillance
	Type 3 (2 - 20 Mbit/s)	NYI	HD video messaging and video clip	HD Video phone and conference big size file transfer	IPTV (drama) P2P download Network game Video conference E-Health appl.	IPTV (movie) P2P download Network game Video conference E-Health appl.
	Type 4 (bigger than 20 Mbit/s)	NYI	3D-Video messaging	3D-based web contents	3D TV 3D-Telepresence Nuclear research appl.	E-Health appl. Nuclear research appl.

NYI: Not Yet Identified

Y.3042(13)_F7-1

Class 0 Class 1 Class 2 Class 3

Some of traffic class 0 requires the highest priority treatment and maintaining of traffic class 1 has significance in public service provision. Traffic class 2 impacts telecommunications including business (new models such as OTT), operation of relevant resources and their management. Traffic class 3 seriously impacts telecommunication operators and ISPs. It is expected that providing leased connectivity, at this stage, would be the best way to support this.

8 Requirements

8.1 High-level requirements

SUN STCRMF has the following high-level requirements for traffic control and resource management:

- It is required to monitor end-to-end IP packets on network elements over fixed/mobile networks;
- It is required to monitor the usage of network element resources;
- It is required to provide mechanisms for controlling traffic and managing resources in order to guarantee fair usage of network resources;
- It is recommended to enable collecting of context information for detailed traffic control and resource management.

8.2 Functional requirements

Functional requirements for SUN STCRMF are divided into six areas: traffic monitoring, traffic analysis, resource monitoring, resource analysis, smart traffic control and resource enforcement. Detailed functional requirements for each area are identified as follows:

- Traffic monitoring related requirements:
  - It is required to collect flow information (e.g., 5 tuples; SrcIP, DestIP, protocol, SrcPort and DestPort);
  - It is required to collect total traffic volume of flow(s);
  - It is recommended to collect traffic context (see Appendix II).
- Traffic analysis related requirements:
  - It is required to classify traffic into flow per user/user-device, volume per flow, volume per user/user-device and volume per service provider;
  - It is required to identify user/user-devices which generate multi-flow;
  - It is required to identify the traffic class of collected traffic volume;
  - It is recommended to generate traffic statistics by time (e.g., real-time, hour, day, week, month);
  - It is recommended to generate traffic statistics by location (e.g., last mile, access, POP, core).
- Resource monitoring related requirements:
  - It is required to collect bandwidth usage information (e.g., used and available bandwidth) per network element;
  - It is required to collect bandwidth usage information per flow in a network element;
  - It is recommended to collect other resource information (e.g., CPU usage, interface utilization, memory usage, storage usage, etc.).
- Resource analysis related requirements:
  - It is required to classify resource usage information per user/user-device or service provider;
  - It is required to identify the overload condition of the network caused by heavy user traffic;
  - It is required to identify the overload condition of the network caused by heavy service traffic;
  - It is required to identify the overload condition of the network caused by heavy signalling traffic;
  - It is required to identify the overload condition of the network caused by surge traffic.
- Smart traffic control related requirements:
  - It is required to control part or all of user flows causing the network overload condition;
  - It can optionally support video pacing (translating), video transcoding, and/or HAS to avoid the network overload condition.
- Resource management related requirements:
  - It is required to support priority-based control (e.g., DSCP marking or QCI channel allocation);
  - It is required to support shaping, queuing, policing, buffer management and admission control.

9 Architecture

9.1 High level architecture

In [ITU-T Y.3041], six capabilities are described to realize SUN. They are context awareness capabilities, content awareness capabilities, programmable capabilities, smart resource management capabilities, autonomic network management capabilities, and ubiquitous capabilities. Of these, the smart resource management capability which is realized with the smart traffic control and resource management function is specified in this Recommendation.

Figure 9-1 shows a high-level architecture of smart traffic control and resource management. Traffic monitoring and analysis are performed at the interfaces of service providers (e.g., CP, VNO, Portal, etc.) and users of mobile or fixed access. In addition, resource monitoring and analysis are also performed over entire networks. Based on the traffic control and resource management as well as context aware capabilities, overload conditions can be controlled. The functions and functional entities of the smart traffic control and resource management are described in clause 9.2.

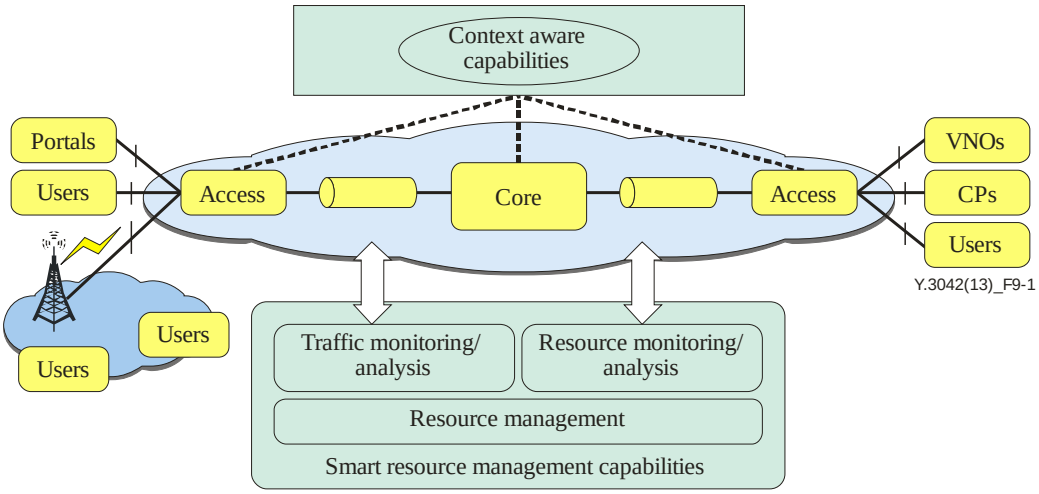

Figure 9-1 – High level architecture of STCRMF

9.2 Functional architecture

This clause specifies the functional architecture with descriptions of functions and functional entities. The STCRMF includes the traffic monitoring and analysis function (TMAF), the resource monitoring and analysis function (RMAF) and the smart traffic and resource control function (STRCF) as illustrated in Figure 9-2. The STCRMF provides interworking reference points with the context awareness functions to get context information.

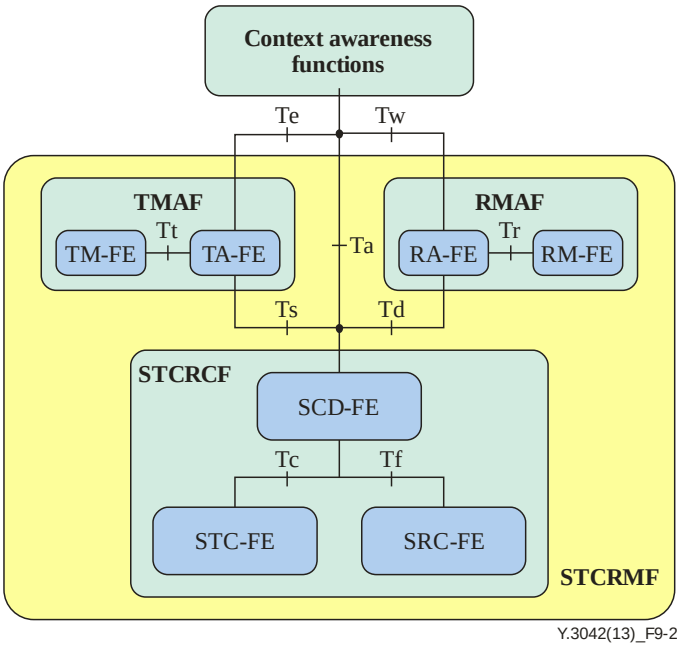

Figure 9-2 – Functional architecture of STCRMF

9.2.1 Traffic monitoring and analysis function (TMAF)

The TMAF monitors and analyses uploading/downloading traffic from/to service providers and users. It consists of traffic monitoring functional entity (TM-FE) and traffic analysis functional entity (TA-FE).

- The TM-FE monitors uploading and downloading traffic at the interfaces of service providers and users respectively. It also communicates with the TA-FE for traffic analysis.
  1) It monitors traffic volume of services to identify which service provider generates heavy service traffic.
  2) It monitors traffic volume of users to identify heavy user traffic.
  3) It monitors traffic volume of users and services to identify heavy signalling traffic.
  4) It monitors traffic volume of users and services to identify surge traffic.
  5) It sends the monitored traffic to TA-FE for traffic analysis.
- The TA-FE analyses and classifies the monitored traffic. For this it also communicates with context awareness functions.
  1) It classifies and analyses monitored traffic per user and service provider based on the predefined policy (e.g., the threshold).
  2) It may generate time-based statistics of classified traffic (e.g., real-time, hour, day, week and month).
  3) For more precise analysis, it may request context information from the context awareness functions.
  4) It sends the analysis results to the SCD-FE.

9.2.2 Resource monitoring and analysis function (RMAF)

The RMAF performs monitoring of network resources (bandwidth usage) as well as analysis of each flow or user. The RMAF includes the resource monitoring functional entity (RM-FE) and the resource analysis functional entity (RA-FE).

- The RM-FE monitors the bandwidth usage of each user or flow. It also communicates with the RA-FE.
  1) It monitors bandwidth usage information per network element.
  2) It monitors bandwidth usage information per flow in a network element.
  3) It monitors other resource information (e.g., CPU usage, memory usage, storage usage, link utilization, etc.).
  4) It informs the RA-FE of the results for resource analysis.
- The RA-FE analyses the monitored resource usage information and identifies the network overload condition. To do this it also communicates with context awareness functions.
  1) It classifies and analyses monitored resource usage information to identify heavy service/user/signalling and surge traffic causing a network overload condition.
  2) It also identifies network failures the information from which can assist in network overload analysis.
  3) For more precise analysis, it requests context information from the context awareness functions.
  4) It sends the analysis results to SCD-FE.

9.2.3 Smart traffic and resource control function (STRCF)

The STRCF determines the optimal control mechanism to manage heavy and surge traffic. It also controls monopolizing users or flows. The STRCF includes the smart correlation and decision functional entity (SCD-FE), the smart traffic control functional entity (STC-FE) and the smart resource control functional entity (SRC-FE).

- The SCD-FE determines the optimal control mechanism by correlation analysis among the status of traffic, resources and context information. It also communicates with context awareness functions, TA-FE and RA-FE.
  1) It receives context information (e.g., user behaviour, device type, service type, location, and content type and time, etc.) from the context awareness functions.
  2) It correlates traffic volume, the resource usage and context information.
  3) It determines which service providers or users generate or receive heavy traffic. In addition, it determines which user or flow monopolizes the network resource, as well as which users or flows are affected.
  4) It determines the optimal control mechanism for smart traffic and resource control.
- The STC-FE controls the uploading/downloading of traffic from/to service providers or users which generate/receive heavy traffic. It communicates with SCD-FE.
  1) It performs application and network layer traffic control (e.g., packet filtering, flow blocking, pacing and transcoding) for the heavy traffic.
  2) It sends a backward warning signal to a service provider or user as a means of indirect traffic control.

- The SRC-FE controls bandwidth usage and resources. It communicates with SCD-FE.
  1) It performs bandwidth reallocation to ensure fair-usage of relevant network resources for every user.
  2) It performs the legacy resource management mechanisms (e.g., marking, policing, shaping, and priority control) to enforce resource control in the event of monopolized network bandwidth usage.

9.3 Reference points

9.3.1 Reference point Ta

The reference point Ta allows the SCD-FE to receive context information needed to determine the optimal control mechanism from context awareness functions. In addition, the SCD-FE can provide notification of the processing result of the control.

The following information flows are exchanged through the Ta reference point.

9.3.1.1 Context information request

It is sent by the SCD-FE to the context awareness functions to request the context information.

9.3.1.2 Context information response

It is sent by the context awareness functions to SCD-FE to confirm the context information request. It contains the requested context information.

9.3.1.3 Context notification

The context notification information flow is sent by SCD-FE to context awareness functions to provide information about the processing result of the control mechanism.

9.3.2 Reference point Te

The reference point Te allows TA-FE to receive the service context information needed for analysis of the monitored traffic from the context awareness functions. The following information flows are exchanged through the Te reference point.

9.3.2.1 Service context request

It is sent by TA-FE to context awareness functions to request the service context information.

9.3.2.2 Service context response

It is sent by the context awareness functions to TA-FE to confirm the service context request. It contains the requested service context information.

9.3.3 Reference point Tw

The reference point Tw allows the RA-FE to receive the user context information needed to monitor and analyse resource usage from the context awareness functions. The detailed user context information is given in Appendix II.

The following information flows are exchanged through the Tw reference point.

9.3.3.1 User context request

It is sent by RA-FE to the context awareness functions to request the service context information.

9.3.3.2 User context response

It is sent by the context awareness functions to RA-FE to confirm the service context request. It contains the requested service context information.

9.3.4 Reference point Tt

The reference point Tt allows TA-FE to receive monitored traffic information for traffic analysis from TM-FE. The traffic monitoring information contains the monitored traffic volume of each service

The following information flows are exchanged through the Tt reference point.

9.3.4.1 Traffic analysis request

It is sent by TM-FE to TA-FE to request an analysis of monitored traffic.

9.3.4.2 Traffic analysis response

It is sent by TA-FE to TM-FE to confirm the traffic analysis request.

9.3.5 Reference point Tr

The reference point Tr allows RA-FE to receive the monitored resource usage information needed for analysis of resource usage from RM-FE. The resource usage monitoring information contains the bandwidth usage of each user (or flow).

The following information flows are exchanged through the Tr reference point.

9.3.5.1 Resource analysis request

It is sent by RM-FE to RA-FE to request an analysis of monitored resource usage.

9.3.5.2 Resource analysis response

It is sent by RA-FE to RM-FE to confirm the resource analysis request.

9.3.6 Reference point Ts

The reference point Ts allows SCD-FE to receive the control mechanism decision information needed to determine the optimal control mechanism from TA-FE. The control mechanism decision information contains the service type and traffic volume per unit time of each service and other pertinent information.

The following information flows are exchanged through the Ts reference point.

9.3.6.1 Smart control decision request

It is sent by TA-FE to SCD-FE to request the optimal control mechanism decision.

9.3.6.2 Smart control decision response

It is sent by SCD-FE to TA-FE to confirm the smart control decision request.

9.3.7 Reference point Td

The reference point Td allows SCD-FE to receive the control mechanism decision information needed to determine the optimal control mechanism from RA-FE. The control mechanism decision information contains the user type and bandwidth usage per unit time of each user as well as other pertinent information.

The following information flows are exchanged through the Td reference point.

9.3.7.1 Smart control decision request

It is sent by RA-FE to SCD-FE to request the optimal control mechanism decision.

9.3.7.2 Smart control decision response

It is sent by SCD-FE to RA-FE to confirm the smart control decision request.

9.3.8 Reference point Tc

The reference point Tc allows STC-FE to receive the traffic control request information needed for traffic control enforcement from SCD-FE. The traffic control request information contains the selected control mechanism and control data, as well as other pertinent information.

The following information flows are exchanged through the Tc reference point.

9.3.8.1 Traffic control request

It is sent by SCD-FE to STC-FE to request the traffic control enforcement.

9.3.8.2 Traffic control response

It is sent by STC-FE to SCD-FE to confirm the traffic control enforcement request.

9.3.9 Reference point Tf

The reference point Tf allows SRC-FE to receive the resource control request information needed for resource control enforcement from SCD-FE. The resource control request information contains the selected control mechanism and control data and other necessary information.

The following information flows are exchanged through the Tf reference point.

9.3.9.1 Resource control request

It is sent by SCD-FE to SRC-FE to request the resource control enforcement.

9.3.9.2 Resource control response

It is sent by SRC-FE to SCD-FE to confirm the resource control enforcement request.

10 Mechanisms for smart traffic control and resource management

10.1 Mechanism based on data cap

Procedures of the traffic control and resource management mechanism based on a data cap are described below:

1) The TM-FE monitors the total traffic volumes of the heavy users. (See Appendix I).
2) Simultaneously, the RM-FE monitors the bandwidth and other resource usage of the network elements associated with the users (or flows) of those services.
3) The TA-FE analyses and identifies the total level of the traffic volume as to whether it reaches the limit in the pre-defined policy (e.g., 70%, 90% or 100% of the data cap). It then sends the analysis result to SCD-FE. Context information from context awareness functions can be optionally used to assist more precise analysis.
4) The RA-FE also analyses the monitored resource information. It identifies the usage level of network resources as to whether it degrades the services to other users (or flows). Context information from context awareness functions can be optionally used to assist more precise analysis.
5) The SCD-FE receives the analysis results by the TA-FE and RA-FE. It correlates the traffic volumes and the resource usage and it determines the data-cap-based traffic control and resource management mechanism. Context information from context awareness functions can be optionally used to assist more precise analysis.

6) The SRC-FE enforces the decision according to a pre-defined policy as follows:

- In the case of 70% of the data cap, notify the user and apply a 50% rate limit of available resources.
- In the case of 90% of the data cap, notify the user and apply a 30% rate limit of available resources.
- In the case of 100% of the data cap, notify the user and apply a 1% rate limit (minimum 1 Mbit/s) of available resources.

Figure 10-1 below shows the procedures for this mechanism.

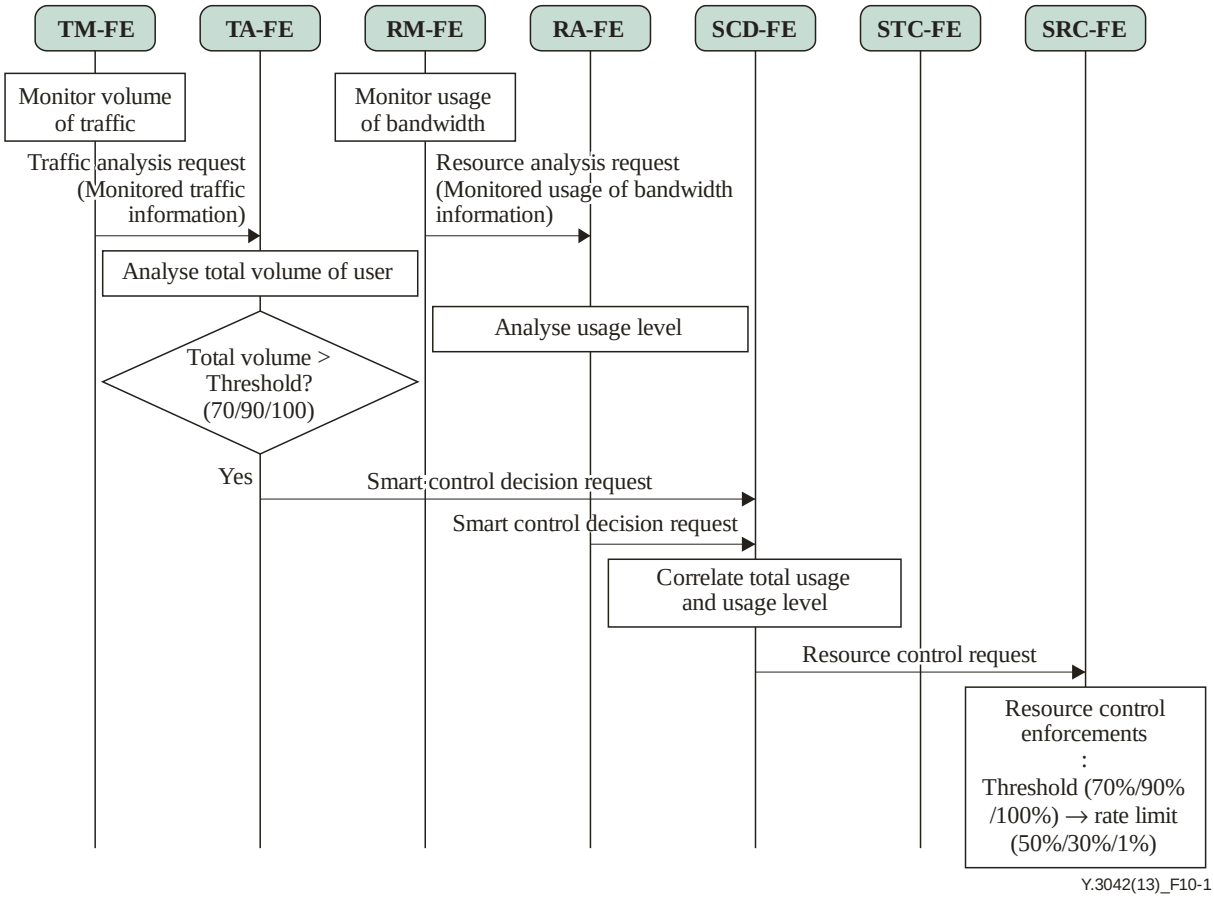

Figure 10-1 – Procedures for the data cap based mechanism

10.2 Mechanism for heavy service traffic

The procedures of the traffic control and resource management mechanism for heavy service traffic are described below:

1) The TM-FE monitors the traffic volume and sends a traffic analysis request containing the monitored traffic information to the TA-FE.
2) The TA-FE analyses the traffic volume for each service.
3) The TA-FE checks whether the traffic volume per unit hour is above the threshold. If it is, the TA-FE sends a smart control decision request to the SCD-FE.
4) Simultaneously, the RM-FE monitors bandwidth usage of network resources and sends a resource analysis request containing the monitored usage information to the RA-FE.
5) The RA-FE analyses the usage level and identifies heavy service candidates.
6) The RA-FE checks whether the resource usage is above the threshold. If it is, the RA-FE sends a smart control decision request to the SCD-FE.

7) The SCD-FE then correlates the traffic volume and resource usage. The SCD-FE identifies a heavy service.

8) The SCD-FE sends a traffic control request to the STC-FE. The STC-FE performs traffic control enforcement (e.g., flow discard (of multi-flows) and flow shaping).

9) Simultaneously, the SCD-FE sends a resource control request to the SRC-FE. The SRC-FE performs resource control enforcement (e.g., policing, shaping and priority control).

Figure 10-2 shows the procedures for this mechanism.

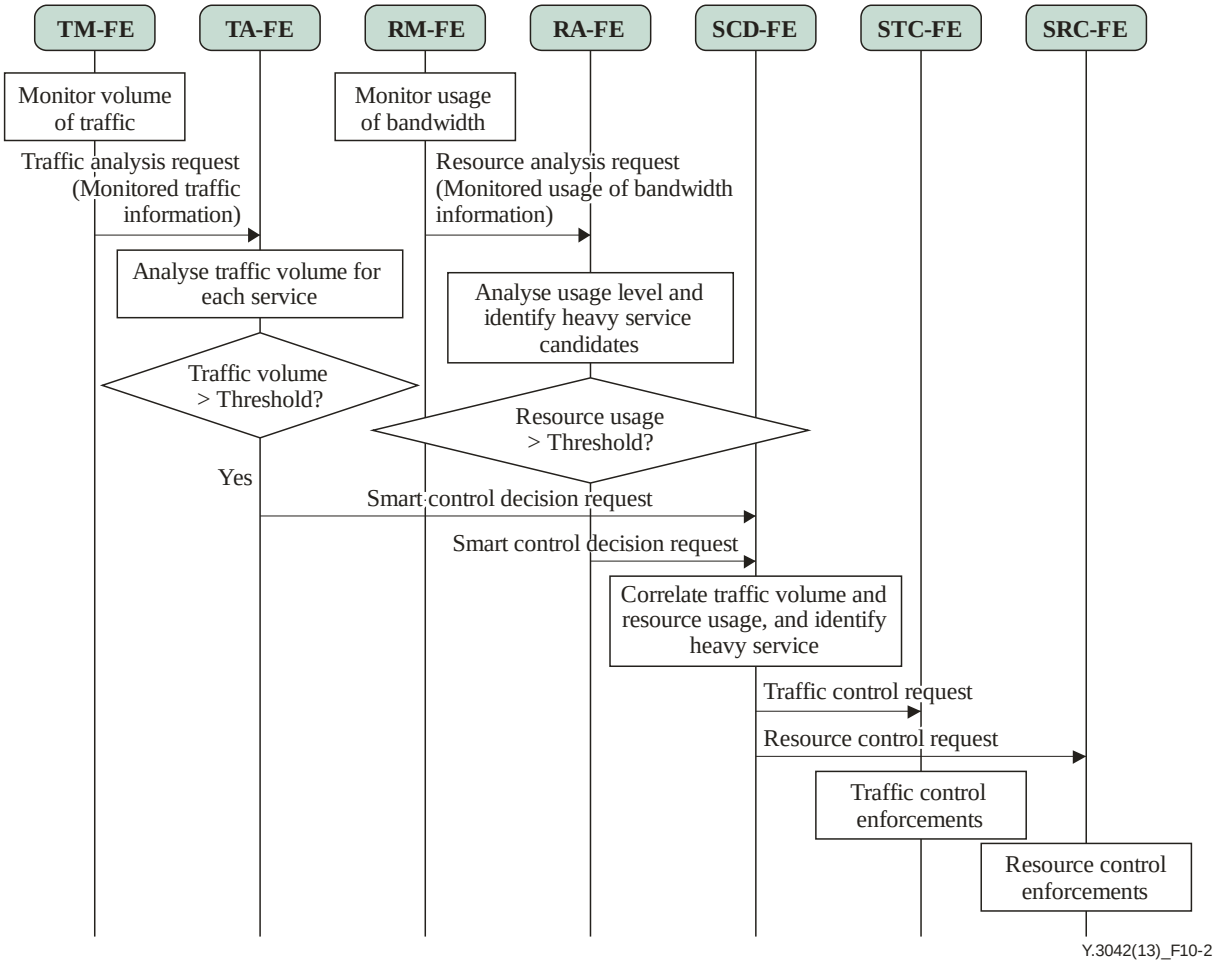

Figure 10-2 – Procedures for the heavy service traffic control mechanism

10.3 Mechanism for heavy signalling traffic

Procedures of the traffic control and resource management traffic for heavy signalling traffic are described below:

1) The TM-FE measures signalling packets.

2) The TA-FE analyses the signalling frequency.

3) If the frequency exceeds the threshold, the SCD-FE correlates the monitored traffic volume and resource usage. It makes optimal traffic control and resource management decisions against heavy signalling traffic. Context information from context awareness functions can be optionally used to assist more precise analysis.

4) The STC-FE notifies the client and server to adjust the signalling frequency. It also controls packets (e.g., the blocking of packets).

Figure 10-3 below shows the procedures for this mechanism.

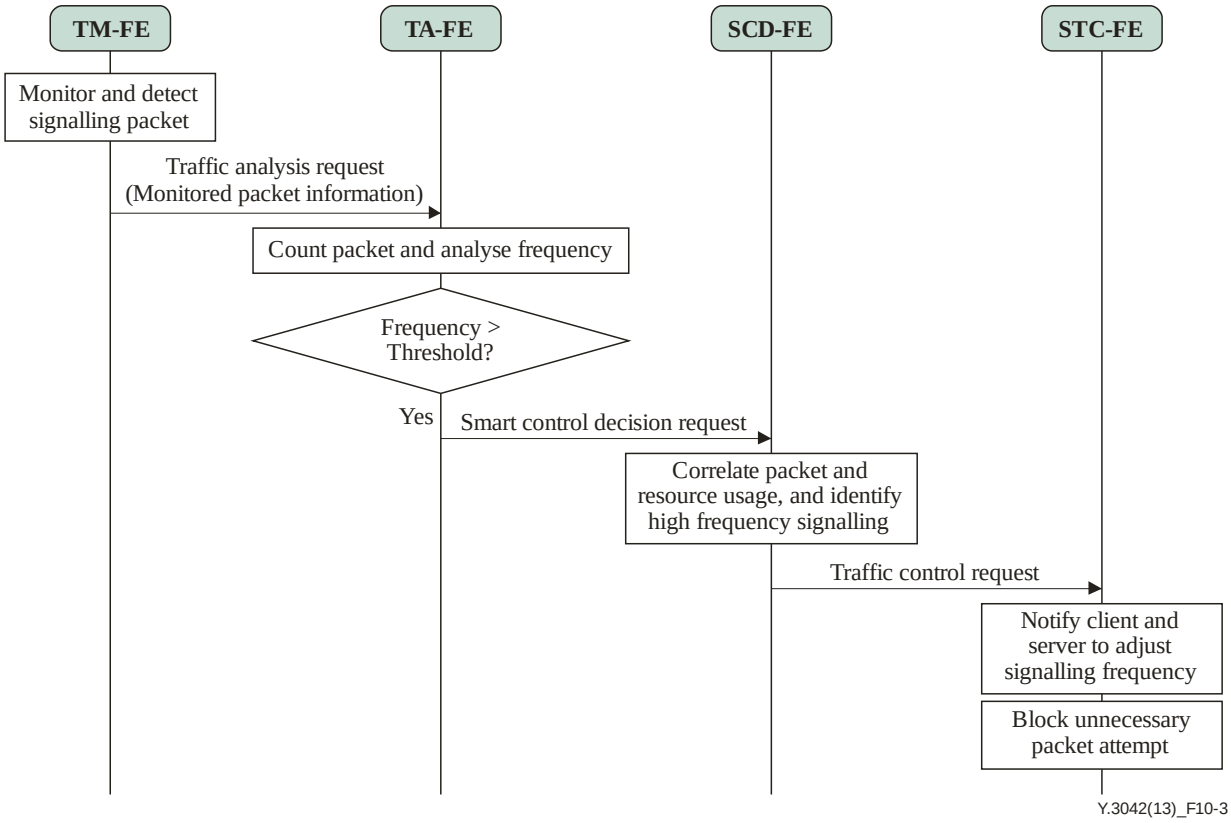

Figure 10-3 – Procedures for heavy signalling traffic

10.4 Mechanism for heavy user traffic

The procedures of the traffic control and resource management mechanism for the heavy user traffic are described below:

1) The TM-FE monitors the traffic volume and sends a traffic analysis request containing the monitored traffic information to the TA-FE.
2) The TA-FE analyses the traffic volume of each service.
3) The TA-FE checks whether the traffic volume per unit hour is above the threshold. If it is, the TA-FE sends a smart control decision request to the SCD-FE.
4) Simultaneously, the RM-FE monitors the bandwidth usage of network resources and sends a resource analysis request containing the monitored usage information to the RA-FE.
5) The RA-FE analyses the usage level and identifies heavy user candidates.
6) The RA-FE checks whether the resource usage is above the threshold. If it is, the RA-FE sends a smart control decision request to the SCD-FE.
7) The SCD-FE then correlates the traffic volume and resource usage and identifies a heavy user.
8) The SCD-FE sends a traffic control request to the STC-FE. The STC-FE performs traffic control enforcement (e.g., flow discard (of multi-flows) and flow shaping).
9) Simultaneously, the SCD-FE sends a resource control request to the SRC-FE. The SRC-FE performs resource control enforcement (e.g., policing, shaping and priority control).

Figure 10-4 shows the procedures for this mechanism.

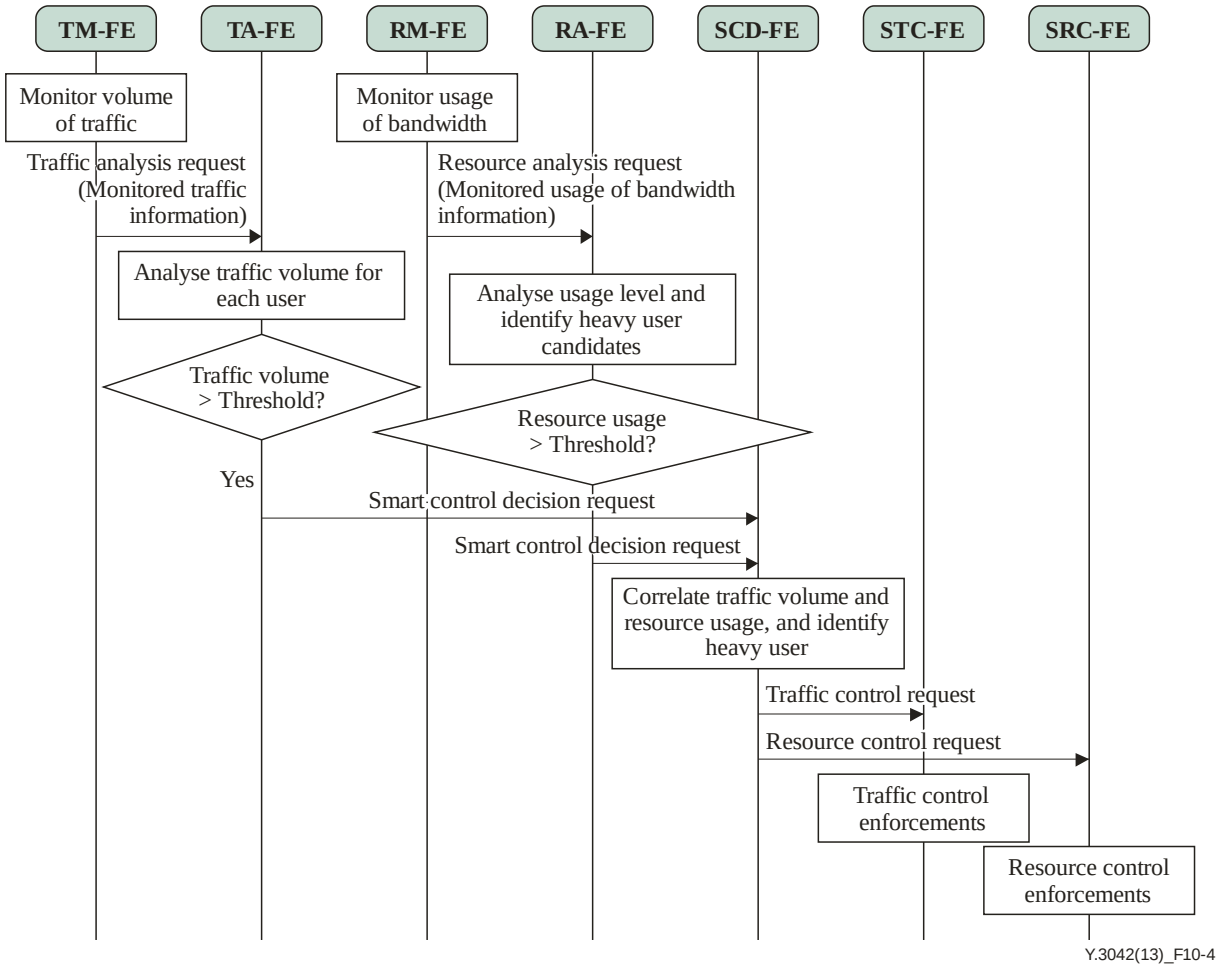

Figure 10-4 – Procedures of the heavy user traffic control mechanism

10.5 Mechanism for surge traffic

The procedures of the traffic control and resource management mechanism for the surge traffic are described below.

1) The TM-FE monitors traffic volume and sends traffic analysis requests containing monitored traffic information to TA-FE.

2) The TA-FE analyses service characteristics (i.e., time/non-time-sensitive) and traffic volume for each service.

3) The TA-FE checks whether the total traffic volume per unit hour is exceeding the threshold. If it is, the TA-FE sends a smart control decision request to the SCD-FE.

4) Simultaneously, the RM-FE monitors the bandwidth usage of network resources and sends a resource analysis request pertaining to the monitored usage information to the RA-FE.

5) The RA-FE analyses the service characteristics (i.e., time/non-time-sensitive) and usage level.

6) The RA-FE checks whether the total resource usage is exceeding the threshold. If it is, the RA-FE sends a smart control decision request to the SCD-FE.

7) Then, the SCD-FE correlates the service characteristics (i.e., time/non-time-sensitive), traffic volume and resource usage. The SCD-FE determines control methods according to control targets.

8) The SCD-FE sends a traffic control request to the STC-FE. The STC-FE performs traffic control enforcement (e.g., flow discard (of multi-flows) and/or flow shaping).

9) Simultaneously, the SCD-FE sends a resource control request to the SRC-FE. The SRC-FE performs resource control enforcement (e.g., policing, shaping, and priority control).

Figure 10-5 shows the procedures for this mechanism.

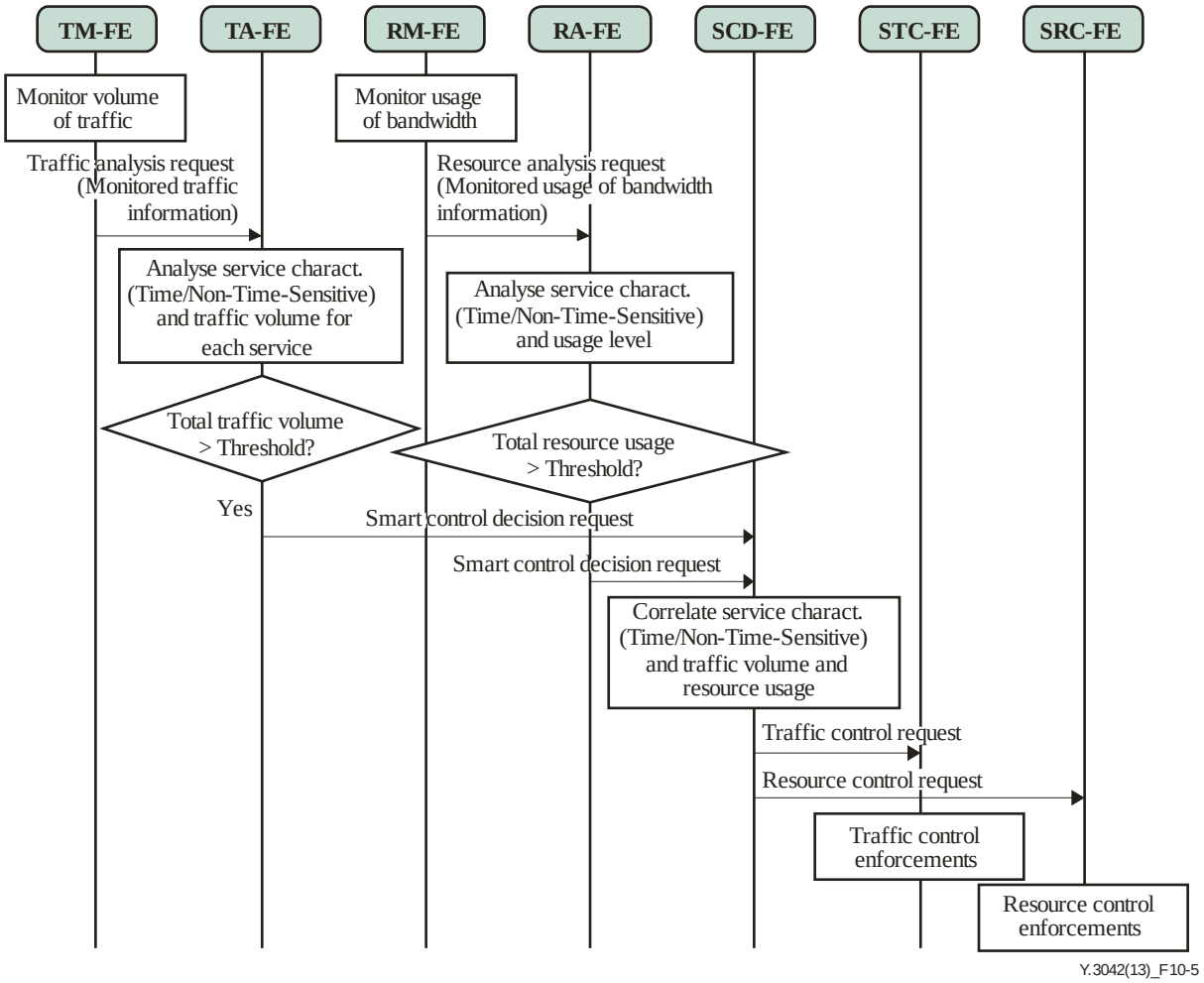

Figure 10-5 – Procedures of the surge traffic control mechanism

10.6 Mechanism for over-sized traffic

Procedures of the traffic control and resource management mechanism for oversized traffic are described below:

1) The TM-FE monitors services and their traffic volume.

2) Simultaneously, the RM-FE monitors the bandwidth and other resource usage of network elements associated with the users (or flows) for those services.

3) The TA-FE analyses the monitored traffic and classifies it according to the predefined traffic volume, the frequency of usage and the device type. Context information from context awareness functions can be optionally used to assist more precise analysis.

4) The RA-FE also analyses monitored traffic. It identifies the usage level of the network resources and whether it degrades services of other users (or flows). Context information from context awareness functions can be optionally used to assist more precise analysis.

5) The SCD-FE receives the analysis results of the TA-FE and RA-FE. It correlates the traffic volume and the resource usage, and makes optimal traffic control and resource management decisions against the oversized traffic. Context information from context awareness functions can be optionally used to assist more precise analysis as well.

6) The STC-FE enforces the decided traffic control decision based on the pre-defined policies or mechanisms (e.g., threshold, packet filtering or transcoding).

7) The SRC-FE enforces bandwidth and resource management decisions based on the pre-defined policies or mechanisms (e.g., threshold, policing, shaping or priority control).

Figure 10-6 below shows the procedures for this mechanism.

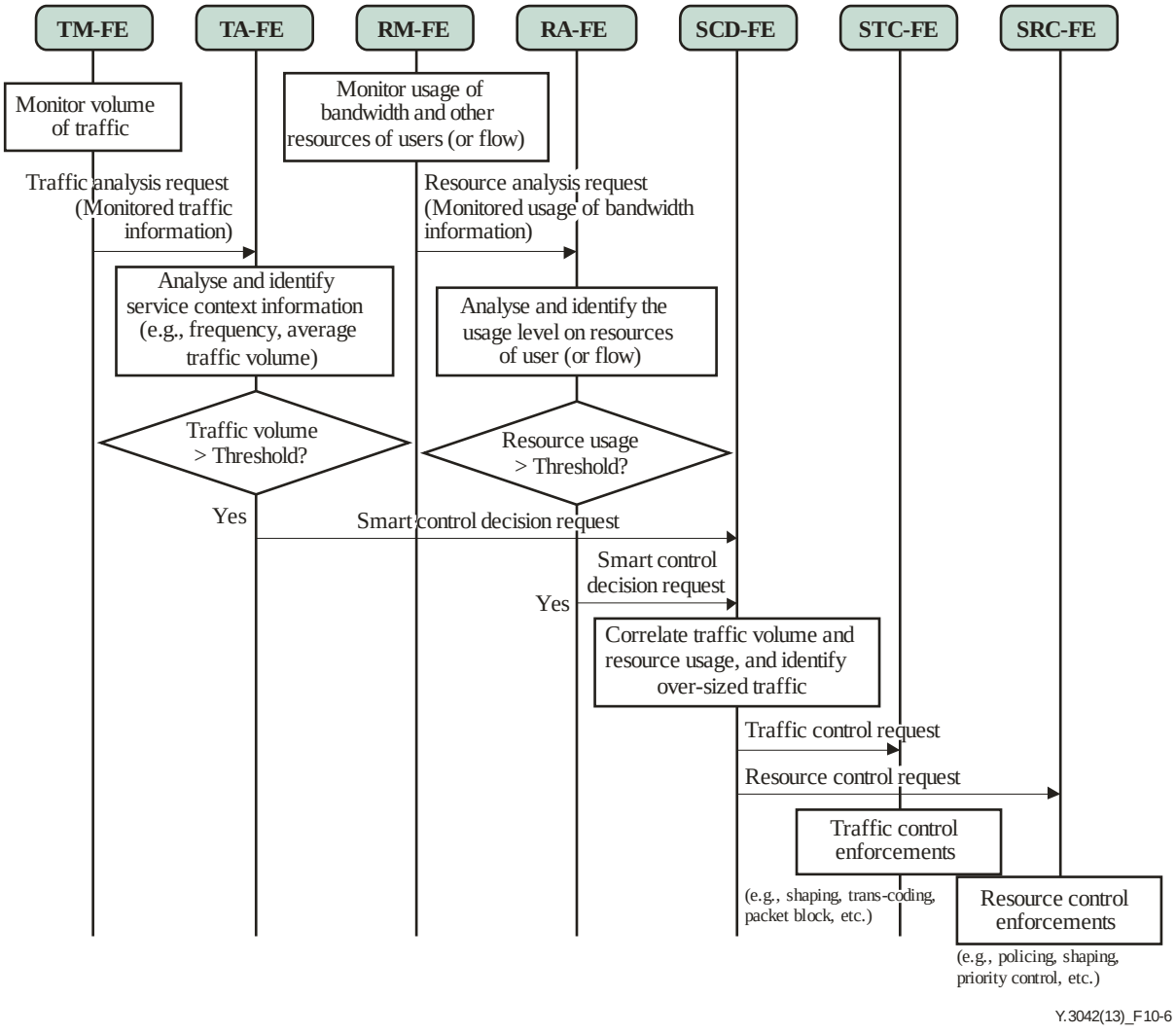

Figure 10-6 – Procedures for over-sized traffic

10.7 Mechanism for busy-hour traffic

The total volume of traffic in an IP network fluctuates continuously over time (time/day/week/month/year) [ITU-T E.600]. In particular, during a specific time period when there are simultaneously multiple users, the network can be overloaded. Traffic control and resource management in such a busy-hour period can alleviate the overloaded condition.

A mechanism for busy-hour traffic control and resource management for non-time-sensitive services can be used.

Procedures of the traffic control and resource management mechanism for busy-hour traffic are described below:

1) The TM-FE monitors the services and their traffic volumes.

2) Simultaneously, the RM-FE monitors the bandwidth and other resource usage of the network elements associated with users (or flows) for these services.

3) The TA-FE analyses the service characteristics (time-sensitive/non-time-sensitive) and the monitored traffic volume of each service. Context information from context awareness functions can be optionally used to assist in more precise analysis.

4) The RA-FE analyses the service characteristics (time-sensitive/non-time sensitive) and the monitored traffic generated by the users (or flows). It identifies the usage level of network resources as to whether it degrades the services to other users (or flows). Context information from context awareness functions can be optionally used to assist in more precise analysis.

5) The SCD-FE receives the analysis results by the TA-FE and RA-FE. It correlates service characteristics, busy-hour information and resource usage. In addition, it makes optimal traffic control and resource management decisions. Context information from context awareness functions can be optionally used to assist in more precise analysis.

6) The STC-FE enforces the decided traffic control decision based on pre-defined mechanisms (e.g., packet filtering or transcoding).

7) The SRC-FE enforces the decided bandwidth and resource management decision based on pre-defined mechanisms (e.g., policing, shaping and priority control).

Figure 10-7 below shows the procedures for this mechanism.

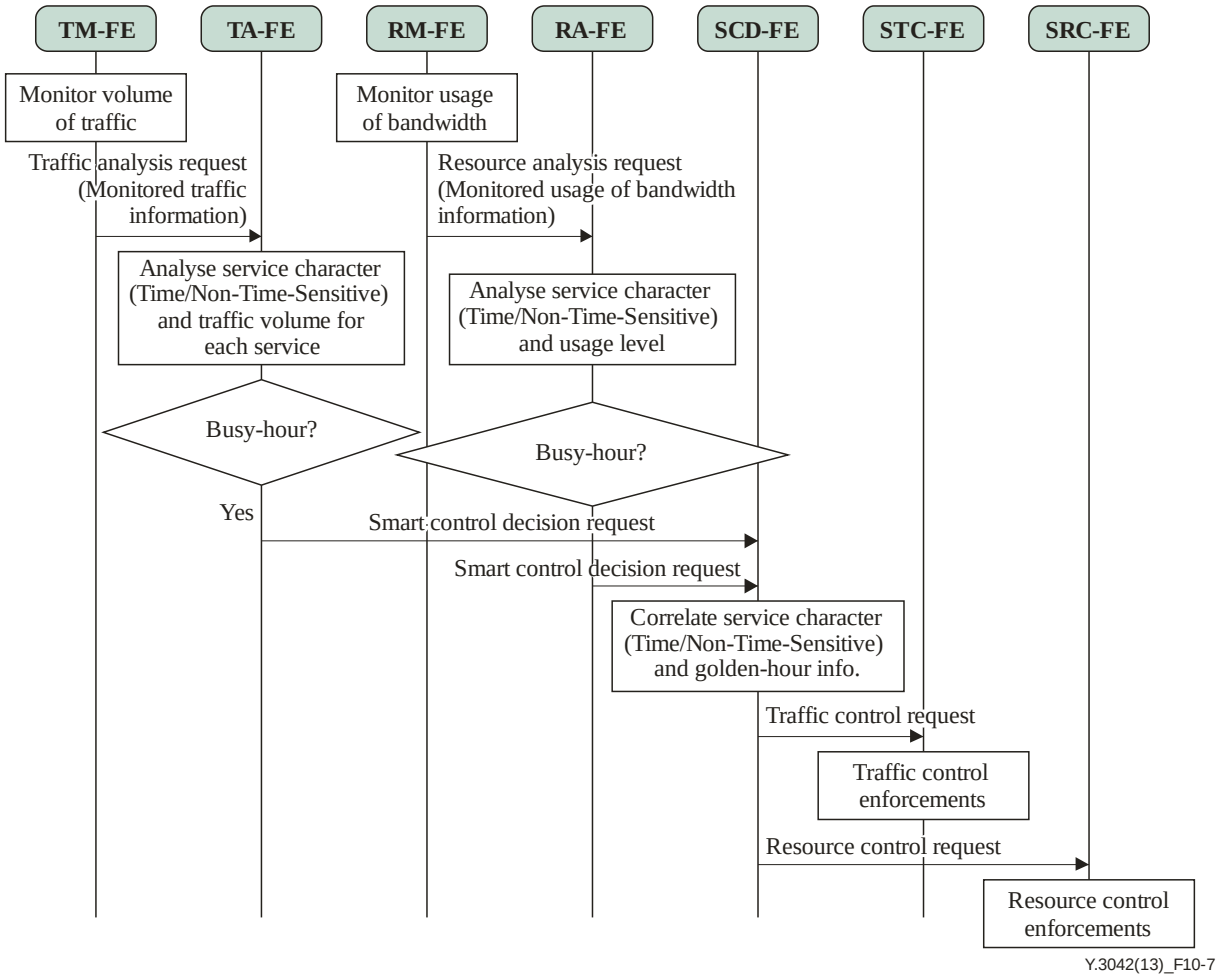

Figure 10-7 – Procedures for busy-hour traffic

10.8 Mechanism based on a list

P2P services may cause bandwidth monopolization and service quality degradation. For effective control of such services, a list-based traffic control and resource management mechanism can be used.

A list-based traffic control and resource management mechanism creates a list of targeted heavy users and controls particular traffic based on the list. The following are two examples of this mechanism specifically for P2P CDN.

– Limit the number of peers connecting to P2P networks and limit the upload speed: limit the maximum number (e.g., 5~10) of peers per type of content, and limit the upload speed (e.g., 500 kbit/s) of P2P for a particular type of content.

– Prevent P2P traffic through the blocking of keep-alive signals (a signal for a client's status information from a server) in a particular P2P CDN: block keep-alive signals among management signals between a P2P server and a client in a hybrid P2P.

In P2P CDN, the procedures for list-based traffic are summarized as below:

1) The TM-FE monitors traffic based on the destination IP address of packets.

2) The TA-FE checks whether the detected packet has a destination IP address that is on the list.

3) If the packet has a destination IP address on the list, the TA-FE further checks whether the packet is a keep-alive packet.

4) If the packet is a keep-alive packet, it sends an analysis request to the SCD-FE. The SCD-FE correlates the packets and their resource usage and determines the optimal traffic control and resource management mechanism. Context information from context awareness functions can be optionally used to assist in more precise analysis.

5) The STC-FE enforces decisions made by the SCD-FE to the target packets (e.g., the blocking of packets).

Figure 10-8 below shows the procedures for this case.

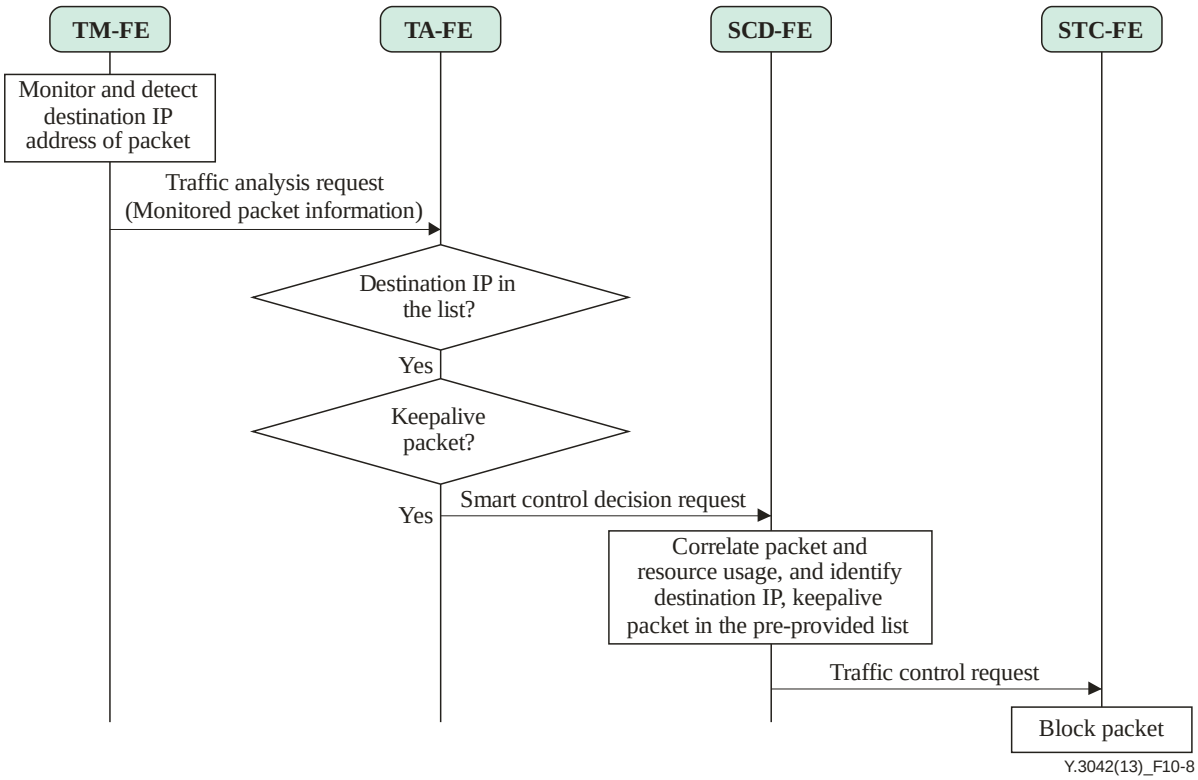

Figure 10-8 – Procedures for list-based traffic in P2P CDN

11 Security consideration

SUN is recognized as an enhancement of IP-based networks. Thus, this Recommendation aligns with the security requirements in [ITU-T Y.2701].

Appendix I

Data explosion and QoS degradation

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

I.1 Data explosion caused by a small number of users in fixed and mobile networks

The introduction of smartphones further increased data traffic in mobile networks. Emerging smart devices which have more capabilities will further increase the data traffic originating in mobile networks (3G, WiFi and WiMax). Consequently, data traffic is increasing in both mobile and fixed networks. In addition, a significant proportion of network resources are being occupied by a small number of heavy users and this has a big impact that is felt more seriously by general users.

Figures I.1 and I.2 show the traffic monopolization situation in mobile and fixed networks. In the case of fixed networks, 5% of the subscribers generate almost 50% of traffic while 20% of the subscribers generate 95%. This indicates that the relevant network resources are not being used properly for all subscribers, but rather that they are being occupied by just a few subscribers. This is more serious in the case of mobile networks where only 1% of the subscribers generate 45% of traffic and 10% of the subscribers use 95% of the resources, which suggests inappropriate usage of the very limited spectrum resources.

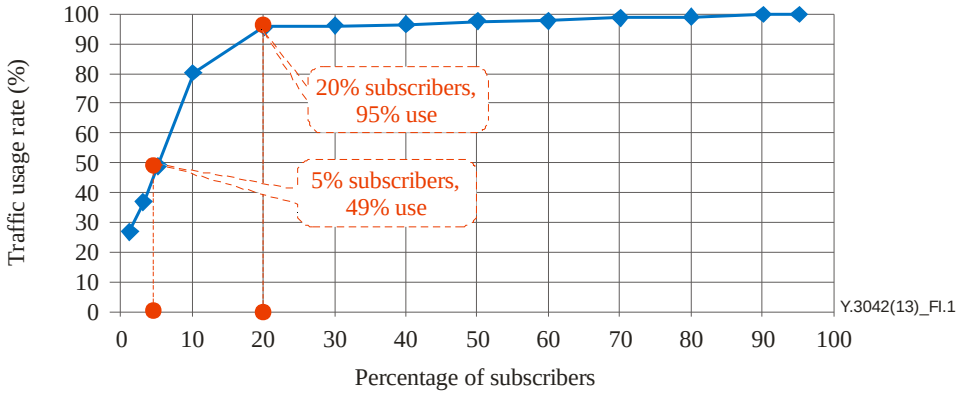

Figure I.1 – Monopolization of traffic in fixed networks, 2010

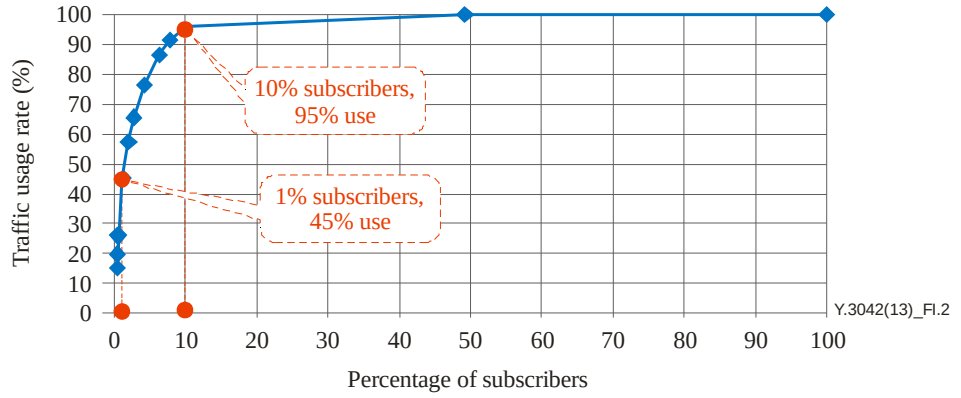

Figure I.2 – Monopolization of traffic in mobile networks, 2010

I.2 QoS degradation for general users caused by a small number of heavy users

In general, a few heavy users seriously impact on the overall quality of services resulting in severe degradation of services for other normal users. Figure I.3 shows severe degradation of QoS caused by one heavy user who occupies almost 97% of the bandwidth using 6 × 100 Mbit/s links. The level of QoS degradation shown in this figure varies from a minimum 29 times to a maximum 265 times. Analysis has shown that normal users who were impacted by such degradation experienced serious difficulties even when performing a simple web search.

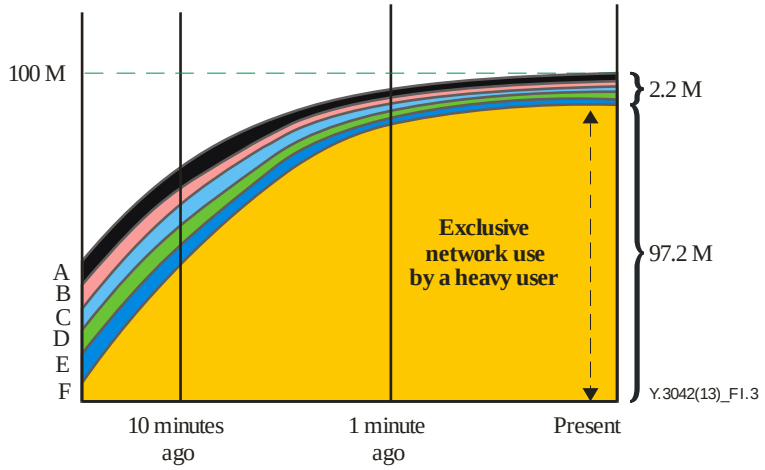

Figure I.3 – Performance decline by heavy and exclusive use

Table I.1 below shows a detailed bandwidth usage situation when the monopolization occurred.

Table I.1 – Performance decline by heavy and exclusive use

User	10 minutes ago	1 minute ago	Present
User A	0.14 Mbit/s	0.15 Mbit/s	0.01 Mbit/s
User B	11.74 Mbit/s	2.45 Mbit/s	0.04 Mbit/s
User C	13.51 Mbit/s	1.86 Mbit/s	1.74 Mbit/s
User D	0.52 Mbit/s	0.35 Mbit/s	0.39 Mbit/s
User E	0.36 Mbit/s	0.06 Mbit/s	0.05 Mbit/s
User F	29.14 Mbit/s	94.85 Mbit/s	97.24 Mbit/s

I.3 Need for smart network management to protect normal user's QoS

After considering the above use cases, it is clear that specific mechanisms carried out in a highly transparent and fair manner are required to protect at least normal user QoS. Smart network management and control capabilities should be able to address these concerns.

Appendix II

Context information examples for STCRMF

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

The STCRMF enables interworking with the context awareness functions to obtain the context information. This information includes user related, device related, service related and contents related data. However, there is not a detailed example for context information in the functional architecture of clause 9.2. Therefore, this appendix provides examples of each context type and information.

Table II.1 shows examples of context information that is used for STCRMF.

Table II.1 – Context information for STCRMF

Context type	Context information
Service context	• Service type: VoIP, audio streaming, video streaming, instant messaging, interactive gaming, Web browsing • Video resolution: SD, HD720p, HD1080p, VGA (Video Graphics Array) • Video encoding (bit rate, bps) • Video frame rate • Codecs: ITU-T H.264, MPEG4 (Motion Picture Experts Group 4)
User Context	• Device type: Smart-phone, smart pad, smart TV, mobile PC • Traffic limit: traffic upper limit • User type: heavy, malicious, general (the result of user behaviour)
Network (Traffic) Context	• Traffic Type: Managed, unmanaged • Network connection type: 3G, LTE, WiMax, WiFi, Fixed • Location, Time • Traffic attribute: Volume, frequency, flow count

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, 2014